

Being Guilty

Being Guilty

*Freedom, Responsibility, and
Conscience in German Philosophy
from Kant to Heidegger*

GUY ELGAT

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

© Oxford University Press 2022

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data

Names: Elgat, Guy, author.

Title: Being guilty : freedom, responsibility, and conscience in German philosophy from Kant to Heidegger / Guy Elgat.

Description: New York, NY, United States of America:

Oxford University Press, [2022] | Includes bibliographical references and index.

Identifiers: LCCN 2021022663 | ISBN 9780197605561 |

ISBN 9780197605585 (epub)

Subjects: LCSH: Philosophy, German. | Guilt. | Conscience. |

Liberty. | Responsibility.

Classification: LCC B2521.E44 2022 | DDC 193—dc23

LC record available at <https://lccn.loc.gov/2021022663>

DOI: 10.1093/oso/9780197605561.001.0001

1 3 5 7 9 8 6 4 2

Printed by Integrated Books International, United States of America

Acknowledgments

Some of the material in this book appeared earlier in article form. An expanded version of “The Individualization of Conscience: What *Daybreak* (9, 10, 544) and *The Gay Science* (117) Tell Us about the Sovereign Individual” (*Inquiry: An Interdisciplinary Journal of Philosophy* 63, no. 1 [2020]: 1–19) has been incorporated into chapter 5, and an expanded version of “Heidegger on Guilt: Reconstructing the Transcendental Argument in Being and Time” (*European Journal of Philosophy* 28, no. 4 [2020]: 911–925) has been incorporated into chapter 6. I thank these journals for granting me the permission to reuse this material. I also wish to thank my family for enabling me to work on this project—especially Valerie, for her unwavering support.

Abbreviations

Works by Nietzsche

- AOM *Assorted Opinions and Maxims*. Translated by John R. Hollingdale. In *Human, All Too Human*. Cambridge: Cambridge University Press, 1986.
- BGE *Beyond Good and Evil*. Translated by Walter Kaufmann. New York: Vintage, 1966.
- BT *The Birth of Tragedy*. Translated by Ronald Speirs. In *The Birth of Tragedy and Other Writings*. Cambridge: Cambridge University Press, 1999.
- D *Daybreak*. Translated by John R. Hollingdale. Cambridge: Cambridge University Press, 1997.
- EH *Ecce Homo*. Translated by Walter Kaufmann. In *On the Genealogy of Morals and Ecce Homo*. New York: Vintage, 1967.
- GM *On the Genealogy of Morals*. Translated by Walter Kaufmann and John R. Hollingdale. In *On the Genealogy of Morals and Ecce Homo*. New York: Vintage, 1967. *On the Genealogy of Morals*. Translated by Carol Diethe. Cambridge: Cambridge University Press, 2006.
- GS *The Gay Science*. Translated by Walter Kaufmann. New York: Vintage, 1974. *The Gay Science*. Translated by Josefine Nauckhoff. Cambridge: Cambridge University Press, 2001.
- HH *Human, All Too Human*. Translated by John R. Hollingdale. Cambridge: Cambridge University Press, 1986.
- KSA *Kritische Studienausgabe*. Edited by G. Colli and M. Montinari. Berlin: De Gruyter, 1999.
- PTAG *Philosophy in the Tragic Age of the Greeks*. Translated by Marianne Cowan. Washington, DC: Gateway, 1998.
- TI *Twilight of the Idols*. Translated by Judith Norman. In *The Anti-Christ, Ecce Homo, Twilight of the Idols: And Other Writings*. Cambridge: Cambridge University Press, 2005.

- WP *The Will to Power*. Translated by Walter Kaufmann and John R. Hollingdale. New York: Vintage, 1968.
- WS *The Wanderer and His Shadow*. Translated by John R. Hollingdale. In *Human, All Too Human*. Cambridge: Cambridge University Press, 1986.

Works by Schopenhauer

- BM *On the Basis of Morality*. Translated by E. F. J. Payne. Indianapolis: Hackett, 1995.
- FW *Prize Essay on the Freedom of the Will*. Translated by E. F. J. Payne. Cambridge: Cambridge University Press, 1999.
- WWR I, II *The World as Will and Representation*, Vols. I and II. Translated by E. F. J. Payne. New York: Dover 1969.

Introduction

I. Why Guilt and Why the History of German Philosophy?

Le problème de la responsabilité n'aurait de sens que si on nous avait consulté avant notre naissance et que nous eussions consenti à être précisément celui que nous sommes.

—Cioran

Now if I had existence from myself, I should have no doubts or wants, and in general nothing would be lacking in me; I should have endowed myself with all the perfections of which I have any idea—in fact I should myself be a God.

—Descartes

Why write an entire book on guilt? What is so special about guilt that justifies a concentrated philosophical occupation with it? Isn't guilt, like so many other states of mind, feelings, or emotions, a singular, transient state which, as such, does not merit the attention of the philosopher, engaged as he or she typically is with the general and lasting? Though guilt—the unpleasant feeling¹ for having

¹ Following the German *Schuldgefühl*, or feeling of guilt, as well as customary English, I presuppose throughout that guilt is a kind of feeling and do not enter the question of whether it should be better understood as an emotion. See Harman and Maley 2019 for a recent defense of the view that guilt should properly be viewed as a feeling rather than an emotion.

done wrong in some sense—is indeed an empirical state that comes and goes and thus appears to be a merely insubstantial affection² (though in some cases it can last for a considerable duration), once one starts to tug at guilt philosophically, an entire web of thoughts and conceptions is shaken along with it, and it is revealed to stand at the very heart of what it is to be human. This is so because guilt stands at the point where a number of lines, so to speak, intersect: morality, or normativity more broadly, since guilt concerns having done *wrong* in some sense; psychology, for it is essentially an *affective* mental state; and metaphysics, for it raises questions about *responsibility* and *freedom*. What, then, can the experience of guilt, the painful sting of the bad conscience, tell us about who we are as human beings, about the kind of being that we possess? What must be true about us for guilt to be possible? How can it be explained or justified?

In this study, I adopt a historical approach in order to think along with six past philosophers about guilt—especially *moral* guilt. Specifically, I focus on the history of German philosophy from the end of the eighteenth century to the beginning of the twentieth century and explore the thought of Immanuel Kant, Friedrich Wilhelm von Schelling, Arthur Schopenhauer, Paul Rée, Friedrich Nietzsche, and Martin Heidegger on guilt, the closely related issues of free will and responsibility, and other issues.

Why, however, specifically concentrate on the history of *German* philosophy? What justifies an occupation with this specific chapter in the history of philosophy? My first reason is this. In contrast to the English or French philosophy of the period I study here, German philosophy was more consistently interested in the question of guilt and in the attempt to make sense of guilt within a broad philosophical framework.³ Though here, too, it should be emphasized, some

² This, of course, is just an approximation. The details of this formula will be filled out in various ways by the philosophers discussed here.

³ Thus, there is scant attention to guilt in the philosophies of David Hume, John Stuart Mill, or Auguste Comte, to name some of the central figures in the English and French traditions of the period I am interested in. The interest of German philosophy in guilt

philosophers had more to say on this question than others, and the recognition of the phenomenon of guilt as philosophically significant grew gradually, from a relatively minor discussion in Kant to a central concept in Heidegger's existential analysis of the human being or "Dasein" (in his terminology). The history of *German* philosophy, then, is the philosophical past that one should explore to deepen one's understanding of the phenomenon.

Why, however, turn to the history of philosophy at all and rummage through the archives of the ideas of the past? Why not conduct a fresh philosophical analysis of the phenomenon and help oneself as needed to the insights of contemporary developmental psychology, neuropsychology, or, say, criminology? I will answer this question by way of giving my second reason for focusing specifically on the history of *German* philosophy, namely, that in the range of thinkers examined here, it presents radically different approaches to the phenomenon of guilt and thus provides an opportunity to survey the phenomenon and its ontological ramifications from different points of view. Specifically, in the study before us, I group together the philosophers discussed under three different headings which correspond to three distinct "traditions" or perspectives.

First, there is what I call the "metaphysical tradition," embodied by Kant, Schelling, and Schopenhauer. Here the phenomenon of guilt is interpreted within the context of a metaphysical view according to which the phenomenal or empirical reality of space and time, governed by a strict causal determinism that leaves no room for freedom and responsibility, should be distinguished from the noumenal or intelligible reality where freedom reigns (or at least can be thought to reign). As a consequence of this metaphysical view, this tradition, on the interpretation defended, subscribes in

might have to do with the great influence Protestantism had on German philosophy, where one's individual responsibility and the personal, unmediated relationship to God is of central importance.

various ways to a very similar idea, namely, that guilt for specific misdeeds, or *empirical guilt*, requires for its justification an act of intelligible freedom by means of which persons, as *causa sui* or causes of themselves, constitute or create themselves. In this act of self-creation or constitution, a person acquires, and can be deemed responsible for, his or her *ontological guilt*, a guilt possessed by virtue of who or what one is. Thus, a closely related similarity of the three thinkers is that they all put forward a version of a *transcendental argument* according to which justified guilt implies intelligible freedom as its condition of possibility. From this metaphysical perspective, guilt, when consistently thought through, reveals the fundamental responsibility each person bears for his or her own empirical life.

Despite these broad underlying similarities, not all three thinkers agree on the specific details. One central difference is that Kant, so I argue, does not wholeheartedly subscribe to the idea of the agent as *causa sui*, though he comes close to endorsing it. In my discussion of Kant's views, I focus especially on his treatment of guilt in his *Metaphysics of Morals* and *Religion within the Boundaries of Mere Reason* and show how empirical guilt, for Kant, presupposes in the final analysis an act of intelligible freedom by means of which the agent endorses his or her fundamental maxim of action and thus constitutes himself or herself as evil or guilty of original sin. I argue that Kant nevertheless ultimately fails to account for the justifiability of guilt, for his position faces a problem of infinite regress: the adoption of the fundamental maxim presupposes a reasoned choice of this maxim, a choice which is itself to be grounded in a yet more fundamental maxim, and so on and so forth.

On the other hand, Schelling and Schopenhauer, I argue, subscribe to the *causa sui* view more explicitly, which enables them to address the problem of infinite regress: the regress is blocked since the human being's self-creation or constitution is grounded in his or her own being or nature. The two thinkers differ, however,

with respect to the metaphysical picture in which they situate their understanding of guilt. Thus, Schelling, in his *Philosophical Investigations into the Essence of Human Freedom*, explains the justifiability of empirical guilt as conditioned by the human being's free act of self-creation whereby it is decided which of the two basic metaphysical principles—that of creaturely self-will or that of universal will, darkness or light—will dominate the other and thus govern the person's being. It is this act of self-creation by virtue of which the agent burdens himself or herself with the ontological guilt that makes empirical guilt possible. Schopenhauer, in his *The World as Will and Representation, On the Basis of Morality*, and *Prize Essay on the Freedom of the Will*, distinguishes three interrelated kinds of guilt but explains personal empirical guilt as conditioned by an intelligible choice of one's intelligible character, in itself an expression of the one underlying will—the basic metaphysical principle in his ontology. It is in the choice of one's intelligible character that one's personal, ontological guilt lies. The first three chapters are devoted to Kant, Schelling, and Schopenhauer, respectively.

The second tradition I identify, the “naturalistic tradition,” is examined through the thought of Rée and Nietzsche. From this perspective, the human being is perceived to fall squarely and fully within the realm of space and time. As such, the human being has no home in a second, noumenal, or intelligible realm and thus has no access to the kind of freedom the metaphysical tradition argues for. This second tradition thus criticizes the metaphysical approach and offers a naturalistic, that is, a psychological-anthropological explanation of guilt. Here the ideas of the person as *causa sui* and of ontological guilt are rejected, and empirical guilt, as well as the false belief in the human being's ontological guilt, are shown to be a result of the working of certain psychological, culturally embedded, and historically developing forces. Importantly, according to this perspective, empirical guilt, while it can be explained, is unjustified insofar as it presupposes the belief in free will or the conception of the person as *causa sui*, which the naturalists examined here reject.

Despite these significant commonalities, Rée's and Nietzsche's approaches diverge.

Rée, in his *The Origin of the Moral Sensations*, takes Darwinian natural selection as his starting point and adopts a broadly associationistic approach to the explanation of the experience of guilt. Specifically, Rée posits the existence of an egoistic and a non-egoistic drive and explains guilt as resulting from the moral disvaluation of egoistically motivated actions that were originally denounced for their long-term disutility. Once such disvaluation fuses with the notions of free will and responsibility, as well as with an unpleasant affective charge, the psychological stage is set for the experience of guilt. Rée develops his position in his later *The Emergence of Conscience* and *The Illusion of Free Will: Its Causes and Consequences* and comes to hold that guilt involves the thought that one deserves to be *punished* but does not subjectively presuppose the belief in free will.

On the other hand, Nietzsche, in his *On the Genealogy of Morals* and other writings (such as *Human, All Too Human*; *Daybreak*; *The Gay Science*, *Beyond Good and Evil*; and *Twilight of the Idols*), focuses in more depth on the historical-cultural genealogy of certain social and mental structures that Rée presupposes and that make the experience of guilt possible in the first place. I examine Nietzsche's criticism of Schopenhauer and reconstruct in detail his complex genealogy of guilt on the basis of his ideas about the internalization of cruelty, the ethics of custom, the debtor-creditor relationship, the creation of the "sovereign individual," the emergence of the belief in free will, and the notion of Christian guilt. The fourth and fifth chapters are devoted to Rée and Nietzsche, respectively. Here I also consider the manner in which Rée's and Nietzsche's analyses compare to each other. Specifically, I claim that Nietzsche's criticisms of Rée fall short of the mark and that the two thinkers' accounts can be seen to complement each other in certain instances.

Finally, the third point of view I examine is that of existential phenomenology, whose representative here is Heidegger. On the one hand, like the naturalists, Heidegger rejects the view that the human being—*Dasein*—is *causa sui* and possesses some sort of intelligible freedom. And yet, on the other hand, and similarly to the metaphysical tradition, Heidegger considers *Dasein* to be ontologically guilty. Thus, on the basis of his existential analytic of *Dasein*, carried out in his *Being and Time*, he claims that we, as *Dasein*, are *Being-guilty*, that is, guilty as such or in our very being, and this is precisely because we are *not causa sui*. Moreover, in a manner reminiscent of the transcendental argument of the metaphysical tradition, Heidegger argues that the *Being-guilty* of *Dasein* is that which makes our empirical or factual guilt possible. I reconstruct this transcendental argument in detail. In addition, I critically discuss Heidegger's phenomenological analysis of the conscience as well as his notion of "wanting to have a conscience." Chapter 6 is devoted to Heidegger.

As the reader may have already noticed, the examination of the history of German philosophy reveals—and this is one of the main claims of this study—that the different approaches to guilt embodied by the three traditions follow upon each other in a dialectical fashion (in a loose sense). Let me formulate this idea more explicitly. Thus, first, with the metaphysical thinkers, I argue that we find the *thesis* that guilt is justified in the final analysis because the human agent is a free *causa sui* or cause of himself or herself.⁴ In response to this view, we have the naturalistic thinkers express the *antithesis* that guilt is not justified since the human agent is neither free nor *causa sui*—to be a *causa sui* is a logical absurdity, according to Nietzsche. Guilt, according to these thinkers, should instead be naturalistically explained. Finally, with Heidegger, we find a *synthesis* of sorts of the two approaches where, on the one hand, the

⁴ As indicated earlier, however, Kant does not clearly embrace this idea but circles around it without fully endorsing it.

idea of *causa sui* is rejected (with the naturalists), but, on the other, a version of the notion that we are ontologically guilty is retained (with the metaphysicians), and guilt is seen as possible, because for Heidegger, as I explain, a condition of possibility for guilt is precisely that we are Being-guilty (ontologically guilty) and yet *not causa sui*.

This, however, leaves us with an apparent aporia: for Nietzsche, guilt is *impossible* (unjustified) because we are not *causa sui*, while for Heidegger, our *not* being *causa sui* is exactly what makes guilt *possible*. In the concluding chapter, this aporia is resolved, and an alternative conception of guilt that builds on Nietzsche's and Heidegger's insights is sketched: guilt as a form of self-torture is indeed unjustified since we do not possess free will (as Nietzsche holds), but as an expression of our normative commitments, it is perfectly intelligible (as Heidegger holds).

Let me now provide my third, more scholarly reason for focusing on the German philosophers' views on guilt. Although many publications in both philosophy and psychiatry or psychology⁵ have concerned themselves with guilt, and although several studies recently published are devoted to the history of *conscience*,⁶ no sustained examination of the history of the philosophy of guilt, either in general or with special emphasis on the German tradition, has appeared. This study aims to fill this lacuna, not only because of the lack of extant research but also because situating each thinker's ideas within their historical context deepens our grasp of the meaning and significance of their arguments. Thus, though some of the figures studied here have received ample *individual* attention from scholars, thinking about their views from within the specific historical arc which I trace here enables us to appreciate them differently. For example, while Nietzsche's views on guilt have been

⁵ See, e.g., Taylor 1985 for a philosophical treatment, Cokelet and Maley 2019 for a more recent philosophical one, and Katchadourian 2011 for the perspective of a psychiatrist.

⁶ See Creveld 2015 and Sorabji 2018.

subjected to a plethora of interpretations, it is only when placed on the background of his engagement with Schopenhauer's views on guilt that the meaning of some of Nietzsche's claims about guilt, both in his early *Human, All Too Human* and in his later *Twilight of the Idols*, can be more fully grasped. To take another example, it is only when Heidegger's thoughts about guilt are read in the context of the problem of the agent as a *causa sui* that his claim about Dasein's not having power over his being "from the ground up" (*Being and Time*, 330) can be appreciated as a philosophical move that engages an entire chapter in the history of philosophical thinking about guilt.

In all these ways, then, the book serves to alter and deepen our understanding of the concepts at issue, the views of the individual philosophers, and the history of German philosophy's thinking about guilt, freedom, and responsibility.

Before I end this part of the introduction, I believe an explanation is in order: I should excuse the absence of G. W. F. Hegel and Sigmund Freud from this study. With respect to Hegel, I decided to leave out his discussions of bad conscience in *The Phenomenology of Spirit* and *Elements of the Philosophy of Right* mainly because his views specifically and the larger framework in which they are situated do not fit well with the argumentative "narrative" the book constructs. And indeed, none of the philosophers I examine refers to him or responds to his ideas, at least not explicitly.⁷ With respect to Freud, my first and simplistic justification is that despite his influence on twentieth-century philosophers (especially those belonging to the continental tradition), he himself was no philosopher and so does not, strictly speaking, belong to the history of philosophy. Second, and more substantially—though I will not defend this perhaps controversial claim here—I believe that despite his conceptual innovations, also deployed in his treatment of guilt (e.g.,

⁷ Heidegger, though, makes a passing reference to Hegel in a footnote to div. 2, chap. 2. See *Being and Time*, 495n6.

id, ego, superego), his account adds little of essence to the other naturalistic views I examine here, those of Rée and Nietzsche.⁸

II. A Few Conceptual Clarifications

Though I end the book with a short sketch of my own ideas on the topic, I abstain from providing a thorough analysis of guilt and an account of how it differs from shame, regret, and remorse.⁹ This is one of the privileges of the historian of philosophy who can observe the battleground from afar and report on the various skirmishes without entering the fray and risking his or her own skin (though reportage carries its own risks!). Nevertheless, a number of initial conceptual clarifications are in order so as to introduce the target topic more clearly. First, what I will be mostly interested in here is the various philosophers' view of the first-person state or experience or feeling of guilt, its nature and conditions (whether transcendental or natural, i.e., psychological-historical). Thus, I will not be concerned here with guilt in the legal sense, where one can be held guilty for breaking the law, say. I will also not be concerned for the most part with objective, third-party attributions of guilt—whether legal or not—by others. Relatedly, I will regard the consciousness and feeling of guilt, when justified, as implying that one is guilty but will allow for the possibility that one can (rightly) be judged guilty by others without experiencing guilt oneself.¹⁰ Furthermore, while focus will be placed on *moral* guilt, I will not assume that guilt can *only* be had and felt with respect to moral transgressions, and, indeed, my discussion of Nietzsche and Heidegger will assume that guilt extends beyond moral guilt.

⁸ For a couple of views that argue for some crucial differences between Nietzsche and Freud, see Miyasaki 2010 and Reginster in an unpublished manuscript.

⁹ See Taylor 1985, Williams 1993b, and Wallace 2017 for some discussion of these distinctions.

¹⁰ I will not argue for this possibility but merely assume it for the sake of my discussion.

In general, I will distinguish in what follows between two basic kinds of guilt, already alluded to above. First, there is *empirical* or *factual* guilt, which an agent incurs by being responsible for an immoral action performed at a specific time and place. This guilt does not make up our very being but is contingent upon what we do. Empirical guilt, furthermore, can be *experienced* by the guilty subject once he or she becomes aware of the wrong nature of the action performed. Here, in addition, we can distinguish (a) retrospective guilt and (b) prospective guilt. By the first I mean the guilt one acquires and can experience subsequently to one's having done something wrong. By the second I mean the guilt one can experience before one acts. Here one feels guilt in advance when merely contemplating an action one knows is impermissible or wrong in some sense. I will mostly put this second kind of empirical guilt aside.

Second is *ontological* guilt; this refers to a guilt that we carry not because of what we *do* (empirically) but because of who or what we *are*. It refers to the ethical imperfection that characterizes our being—whether as a species or individually—and is seen by some of the thinkers examined here as serving to justify and explain our empirical transgressions. Even though, in contrast to empirical guilt, ontological guilt is not acquired as a result of any specific *empirical* action we perform, for some of the thinkers considered here, we are nevertheless responsible for it and consequently *guilty* for it, for we acquire it by means of a free, *intelligible deed*. One extreme expression of the idea of ontological guilt is the thought that the human being is evil or morally corrupt in his or her very being, the religious idea of original sin. As we shall see, Kant, Schelling, Schopenhauer, Nietzsche, and Heidegger interpret the idea of ontological guilt in their own various ways.

Another important distinction to make is that between guilt and *conscience*. The conscience, we may say, is a capacity to know right from wrong (in the moral sense) and as such alerts us to what we should do in certain critical situations. The conscience and

the feeling of guilt are therefore connected in that the experience of guilt, insofar as it presupposes a recognition that one has done wrong, involves knowledge of that wrong, a knowledge that stems from the moral discriminations of the conscience. Following some of the thinkers examined here, I will use the terms “bad conscience” or “pangs of conscience” interchangeably with “feeling of guilt” or “experience of guilt.” While I do address the concept of conscience along the way—in my discussions of Kant and Heidegger mostly—the conscience *as such* is not the main focus of this research.

Other immediately related concepts are those of responsibility and freedom or free will. In everyday parlance, we say that someone is responsible for something or guilty of something and sometimes treat these two expressions as equivalent. But guilt and responsibility should be kept apart. First, while we can say of someone that he or she is guilty of or responsible for something, the former phrase is reserved for what we regard as immoral or unacceptable deeds, while the latter could also be used for deeds we applaud and welcome and thus in itself does not connote any specific value judgment. Furthermore, and more important, justifiably feeling guilty *presupposes* being responsible in some sense. The concept of responsibility itself, however, is ambiguous between mere causal responsibility¹¹ and substantial responsibility (either moral or nonmoral) which grounds blameworthiness or praiseworthiness. Finally, responsibility in this more substantial or robust sense presupposes for several of the thinkers examined here—those belonging to the metaphysical tradition—freedom of the will in some metaphysical sense. One of the central recurring issues in this study concerns the question of whether and how freedom and responsibility can be understood in a way that makes guilt justifiable.

¹¹ This is what Susan Wolf has called responsibility in the “superficial sense”: “When we say that an individual is responsible for an event in the superficial sense, we identify the individual as playing a causal role that, relative to the interests and expectations provided by the context, is of special importance to the explanation of the event” (Wolf 1990, 40).

III. Some Methodological Reflections

Thinking about guilt—and this problem affects all psychological phenomena of a similar nature—raises problems of access. And though these are related, what I have in mind here are not the problems of definition and demarcation already alluded to. What I mean by this is that if we wish to discuss a phenomenon such as guilt, then we must be in a position to access the phenomenon itself in order to obtain accurate cognition of it. Being a first-person phenomenon, however, guilt is not a public event that can be prodded and tested objectively but is mostly given to us in private introspection (which is not to deny, importantly, that guilt can have and typically has external, behavioral manifestations). But what exactly is it that is “given” to us? A sustained reflection on the nature of guilt reveals that a precise answer to this question cannot be easily found, since that which is “given” is always encountered through the medium of various kinds of interpretations of the phenomenon, interpretations that preclude direct cognitive access to the “thing itself” and thus leave us uncertain of its nature. Indeed, such worries could also give rise to the suspicion that there is no such thing as guilt in itself at all and that all we have and all we can consult in our personal experience are just different interpretations of guilt. To be certain, it is possible at least to *try* to dig through the various layers of interpretation that cover over guilt and reach the phenomenon itself in its purity. Indeed, this is what one of the central figures in this book, Nietzsche, can be read as trying to attain. As he puts it in BGE 230:

under such flattering colors and make-ups . . . the basic text of *homo natura* must again be recognized [*wieder heraus erkannt werden muss*]. To translate man back into nature; to become master over the many vain and overly enthusiastic interpretations [*Deutungen*] and connotations that have so far been scrawled and painted over that eternal basic text of *homo natura*.

Whether Nietzsche's naturalization is successful or whether, in contrast, his own view is at the end of the day another interpretation rather than "text" is something with which I will be concerned in the chapter devoted to Nietzsche and in the book's conclusion. But regardless of Nietzsche's specific view on guilt, we face here a second level of complexity: in order to interpret Nietzsche's own texts and so judge whether his attempt at naturalizing guilt is successful, we already have to have at our disposal some understanding of what guilt is; otherwise, we would not be able to tell whether what *he* seems to have before his mind's eye when discussing guilt is what *we* have before ours. We seem to be caught in a hermeneutic circle of sorts: to understand Nietzsche, we have to have some grasp of what guilt is, but to know what guilt is, we have to be able to put aside, with the help of Nietzsche, all the "enthusiastic interpretations" that have been "painted over" it.

Heidegger, famously, believed that this circular predicament is characteristic of the nature of understanding and interpretation as such and is thus unavoidable. The question for him was not how to avoid the circle (an impossibility according to him) but how to get into the circle in the right way (*Being and Time*, 195). But what *is* the right way? Further, since we are dealing in this book with several different thinkers, and not just with Nietzsche, we will have to enter not just one but a number of circles that, moreover, do not lie outside each other but intersect or overlap one another in at least two ways: first, because every thinker's text is to some extent a response to, and thus presupposes its own interpretation of, one of the other philosophers' texts (Nietzsche responding to Rée, Rée responding to Kant and Schopenhauer, etc.); second, because as a reader of all these texts, my interpretation of the one will be inflected by my interpretation of the other, as well as by my interpretation of my own experience of guilt. Further complications ensue when the voices of other commentators are brought into the conversation, and my interpretation of them has to be mapped onto both my reading of the original texts and my own understanding of the phenomenon.

As the reader can appreciate, the task at issue is stupendously complex and can be compared to a dizzying hermeneutical juggling act. Whether I have managed to find a balance between all these different exegetical demands without dropping one of them to the ground is for the reader to decide.

1

Kant

The Timeless Deed That Makes Guilt Possible

Introduction

In this chapter, I examine the thought of Immanuel Kant on guilt and lay the ground for the claim that despite the various differences that separate these three thinkers, at the end of the day, for Kant, Schelling, and Schopenhauer—though less unequivocally in Kant’s case—in order for guilt (empirical or ontological) to be justified, the guilty agent must in some sense be morally responsible for his or her own being; the guilty agent must be his or her own *free* cause or a *causa sui*.

This chapter’s argument, briefly stated, is that empirical guilt at the end of the day could only be justified for Kant if the agent is responsible and consequently guilty for his or her own “original sin” or radical evil (ontological guilt), where this responsibility and guilt imply an intelligible free deed, a position Kant defends in his *Religion within the Boundaries of Mere Reason*. It is here that Kant can be seen to put forward a transcendental argument from (ontological) guilt to intelligible freedom. I conclude the chapter by arguing that Kant, however, does not ultimately succeed in showing why guilt (empirical and ontological) is justified and that even though he can be seen to approach the idea of the subject as *causa sui* that later thinkers endorse, he does not embrace it fully.

I start in the first part with an examination of Kant’s views on the conscience and guilt in his *The Metaphysics of Morals* and then turn in the second part to interrogate Kant’s account of freedom

in the *Metaphysics*. In both parts, I also draw on other texts of Kant's as needed. In the third part, I turn to look at Kant's transcendental argument for transcendental freedom in *Religion within the Boundaries of Mere Reason*. In the conclusion, I raise some final concerns. Importantly, I aim in the following not to trace chronologically the development of Kant's views but to reconstruct his position.

I. Kant on the Conscience in *The Metaphysics of Morals*

Kant does not devote a lot of attention to the concept of conscience or the concept of the bad conscience or guilt, but the little he says is of significance. I will first present his ideas in outline and then proceed to look at them more critically.

In his last central work in practical philosophy, *The Metaphysics of Morals* (1797),¹ Kant lists the conscience as one of the several "moral endowments," along with "moral feeling . . . love of one's neighbor, and respect for oneself (*self-esteem*)"—all of which are "natural predispositions of the mind (*praedispositio*) for being affected by the concept of duty, antecedent predispositions on the side of *feeling*" (MM 6:399). He further claims that the human being is under no obligation to acquire any of these predispositions, precisely because "they lie at the basis of morality, as *subjective* conditions of receptiveness to the concept of duty, not as objective conditions of morality"; rather, it is by virtue of these feelings, Kant says, that the human being "can be put under obligation" in the first place (6:399). Indeed, the very consciousness of these feelings can "only follow from consciousness of a moral law" (6:399). Further, for Kant, the conscience is a psychological *primitive*: it is a basic moral endowment and not something the origin of which can be

¹ Hereafter cited as MM. All references are to the standard pagination.

psychologically accounted for. As he says: “conscience is not something that can be acquired . . . rather, every human being, as a moral being, *has* a conscience within him originally” (6:400).

Kant’s categorization of the conscience as an “antecedent predispositions on the side of *feeling*” (MM 6:399) seems to imply that for him, it is through one’s conscience that one comes to affectively—by means of *feeling*—experience one’s standing vis-à-vis the moral law. Kant however clarifies that conscience is not merely a matter of feeling and claims that “conscience is practical reason holding the human being’s duty before him for his acquittal or condemnation in every case that comes under a law” (6:400). Later, in the section “On the Human Being’s Duty to Himself as His Own Innate Judge,” we learn that things are even more complex: “Consciousness of an *internal court* in the human being (‘before which his thoughts accuse or excuse one another’) is **conscience**” (6:438; emphasis in original).

Kant distinguishes two kinds of cases that can be tried in the “court” of conscience. The first is when the conscience, by means of a representation of that which is morally required in the particular case, merely warns the subject “before he makes a decision” (MM 6:440). This corresponds to what I called prospective guilt. The second case concerns a situation where the action has already been performed and where what I called retrospective guilt can arise. Though the two are similar in essence, in Kant’s presentation, the various legal personae who participate in the “court’s” proceedings only become involved in the second case.² I will focus my discussion on this second case.

² The two cases are similar in essence, I say, because it seems that in both cases, a determination has to be made whether the act at issue—whether already performed or merely considered—conforms to the moral law and is imputable. But this seems to require the working of the various personae Kant mentions in relation to retrospective guilt: a prosecutor and a legal counsel and a judge. The only substantial difference, it seems, is that in the case of prospective guilt, the judge merely warns of a punishment, and in the case of retrospective guilt, the judge gives “effect to his laws” (MM 6:439). I will return to this.

Here, in the case of retrospective guilt, the agent is accused by a prosecutor and enjoys the defense of a “legal adviser” (MM 6:438, footnote), so that at the end of the disputation, a verdict is passed by reason (6:438), acquitting or condemning the accused (6:440), followed by the passing of a sentence by a judge (6:438, footnote). Though Kant mentions all these different personae, he does not explicitly assign a specific mental faculty to each of them or explain in detail the court’s proceedings. If we fill out some of the missing details, the general picture seems to be the following. First, the agent’s practical reason supplies a general representation of the moral law. Then, an act of judgment yields the demands of the law in the agent’s particular case. Then, a comparison is made of the action performed (or omitted)—as the agent perceives it—with the demands of the law in the particular case in order to reach a judgment³ (6:438) as to whether the action committed or omitted conformed with these demands. Finally, and as a consequence of the verdict of this judgment, which is the “conclusion of reason” (6:438), an “internal judge . . . pronounces the sentence of happiness or misery” (6:438, footnote), and thus either a good conscience (in case the action accords with the moral law—a verdict of acquittal) or a bad conscience, that is, guilt (in case the action failed to accord with the moral law—a verdict of condemnation), is experienced.⁴ The former sentence is by nature pleasant, though only in a negative way, as involving the experience of “having escaped the danger of being found punishable” (6:440), while the latter sentence involves a “painful feeling,” the pain of “repentance (*Reue*),” as Kant explains in his discussion of the conscience in the *Critique of*

³ As Thomas Hill emphasizes, “conscience holds up our *acts-as-we-perceive-them* for comparison with the *general-moral-judgments-that-we-accept* . . . in order to see whether we have acted well by our own lights” (Hill 2002, 241).

⁴ Kant defines “judge” or “court” in the following way: “The (natural or moral) person that is authorized to impute with rightful force [i.e., with rightful consequences]” (MM 6:227). I will return to the concept of “imputation.”

Practical Reason (5:98), a discussion that employs very similar legal terminology.

Though this indeed seems to be Kant's general idea, once one starts to examine closely the claim about the conscience being a kind of internal courtroom, questions arise. Let's take a closer look and start with the prosecutor/accuser and the accused. Elaborating on an idea introduced in the text itself, Kant notes in a footnote to the passage on MM 6:438 that a "human being who accuses and judges himself in conscience must think of a dual personality in himself, a doubled self" where "I, the prosecutor and yet the accused as well, am the same *human being* (*numero idem*).¹" Kant clarifies that "the prosecutor" is "the human being as the subject of the moral lawgiving which proceeds from the concept of freedom and in which he is subject to a law that he gives himself . . . [and] is to be regarded as another . . . from the human being as a sensible being endowed with reason." Thus, the prosecutor is the person considered as the "intelligible" I or self, and the accused is the same person considered as a "sensible being," that is, as the sensible I or self, though, as Kant claims, since "there is no theory about the causal relation of the intelligible and the sensible," the distinction is to be maintained "only in a practical respect." What this latter qualification means, it seems, is that though we cannot *cognize* the causal relation between the two selves, that is, *whether* they do in fact interact or *how* they interact in the course of conscience's proceedings, when we represent these proceedings to ourselves, we are necessitated, on pain of reason falling into "self-contradiction" (6:438), to represent to ourselves the situation as if we had a "doubled self" and where these two selves—intelligible and sensible—interact by taking on the judicial roles mentioned (i.e., of prosecutor and accused).

What this "self-contradiction" is Kant does not explain in the footnote itself. But in the body of the text, he makes the following claims:

[C]onscience is peculiar in that, although its business is a business of a human being with himself, one constrained by his reason sees himself constrained to carry it on as at the bidding of *another person*. For the affair here is that of trying a *case* (*causa*) before a court. But to think of a human being who is *accused* by his conscience as *one and the same person* as the judge is an absurd way of representing a court, since then the prosecutor would always lose.—For all duties a human being's conscience will, accordingly, have to think of *someone other* than himself (i.e., other than the human being as such) as the judge of his actions, if conscience is not to be in contradiction with itself. This other may be an actual person or a merely ideal person that reason creates for itself. (MM 6:438)

The worry about a “self-contradiction” seems to be that to think of the courtroom as operated by a *single* person is to make a mockery of the idea of a courtroom, for then the person accused, who would also play the role of prosecutor and judge, would never (or seldom) judge the case on its merits but rather make it so that he or she escape scot-free. To avoid this self-contradiction thus requires *thinking* about the “courtroom” as operated not just by oneself—it requires an introduction of another “person.”

Now, even though Kant claims the other person may be an “actual person” and not necessarily an “ideal person,” he does not go on to elaborate on what he means by this. (Who should this other person be, a judge? A priest? A confidant? An absolute stranger?) Neither does he explain what factors determine under what circumstances it is the real or rather the ideal person who should be involved in one's representation of the internal court of conscience. In addition, the claim that “someone other” than the person prosecuted needs to play the role of judge sits uncomfortably with Kant's claims in the quoted footnote. As we saw, Kant claims there that “a human being who accuses and judges *himself* in conscience must think of a dual personality in himself, a doubled self which, on the one hand has

to stand trembling at the bar of a court that is yet entrusted to him, but which, on the other hand, *itself* administers the office of judge that it holds by innate authority” (emphasis added). In other words, it seems that here in the footnote, and in contrast to the body of the text itself, Kant claims that the role of the judge in conscience may be thought of *not* as “*someone other*,” ideal or not, but as the accused himself, albeit qua his “intelligible” self.⁵

Instead of addressing these issues, Kant proceeds to describe the *ideal* person and claims that “Such an ideal person,” “the authorized judge of conscience,” must be a “scrutinizer of hearts, since the court is set up *within* the human being” (MM 6:439). Further, he must also be thought of as a person “in relation to whom all duties whatsoever are to be regarded as also his commands.” Finally, this “moral being must also have all power (in heaven and on earth) in order to give effect to his laws.” Unsurprisingly, Kant concludes that this kind of being is called “God,” which is why “conscience *must* be thought of as the subjective principle of being accountable to God for all one’s deeds” (6:439; emphasis added), where a “deed”—and I will return to this—is a free action (6:227). It is noteworthy that this claim, in contrast to what we saw Kant holds, in effect rules out the possibility of an *actual* person playing the role of judge needed to make sense of the conscience. In other words, the claim seems to be that in order to avoid the “self-contradiction,” one *must* think of none other than *God* himself as the internal judge in the court of conscience.

Now Kant immediately clarifies that this does not give us any reason to judge that God “*actually exists outside*” of us (MM 6:439), for “the idea [of God] is not given . . . *objectively*, by theoretical

⁵ As a final twist, confusing matters even further, Kant ends the footnote with the following: “When the proceedings are concluded, the internal judge, as a person *having power*, pronounces the sentence of happiness or misery, as the moral results of the deed. Our reason cannot pursue further his power (as ruler of the world) in this function; we can only revere his unconditional *iubeo* [I command] or *veto* [I forbid].” And so it turns out that the judge must be thought of as another person (God) after all, as Kant goes on to argue in the body of the text itself.

reason, but only *subjectively*, by practical reason,” in a manner which merely points “in the direction of thinking of conscientiousness . . . as accountability to a holy being (morally lawgiving reason) distinct from us yet present in our inmost being, and of submitting to the will of this being, as the rule of justice.”

This is a curious piece of reasoning: Kant starts from the idea that the conscience is to be thought of as an internal court of the moral law and ends with the thought that, subjectively, or merely practically speaking, the conscience must be thought as involving a divine presence in us. A crucial link in this reasoning seems to be that “if conscience is not to be in contradiction with itself” (MM 6:438), we must think of the judge in conscience as another person. Does this premise stand up to scrutiny?

Though Kant is surely right to highlight the inner, unpleasant tension that is characteristic of the phenomenology of conscience, especially of bad conscience or guilt, there is no reason to think of the judging instance in conscience as representing or embodying a separate person, whether divine or not. First, the representation of the conscience as an internal court is merely metaphorical: a courtroom where the judge and the accused were the same person would indeed be ridiculous, but there is no reason to think we *must* conceive of the conscience in this metaphorical way. At least the phenomenology of conscience does not force this specific image on us. In response, it might be argued that, metaphors aside, in one’s experience of conscience, judgment, in some sense, *is* passed on oneself, and for this judgment to have any authority with respect to oneself, the judging agency must be distinct or thought of as distinct from the person judged. But there is no reason to think this must be the case, either. One can, for example, judge that one has made an arithmetical mistake in one’s tax calculations and recognize this judgment as authoritative over oneself and before which one must “submit,” without assuming that a separate, distinct person, representing a mathematical judge of some sort, is passing judgment here. Rather, one judges *oneself* to have made a mistake

and feels in addition the bitter regret over one's error as a result of the realization, say, that one is now subject to a federal fine for underpayment. Similarly, there is no reason not to think of the guilty conscience in the same way: one, by the power of one's own pure reason, passes judgment on oneself, proclaims "I have done wrong," accepts the authority of the judgment, and feels the sting of guilt. It is, of course, always possible for motivated reasoning or self-deception to sneak into one's reflections and get one off the hook. But the possibility and reality of such distortions in reflection do not necessitate the inward split Kant mentions as belonging essentially to conscience, not to mention the introduction of a divine "scrutinizer of hearts" (MM 6:439).

Let us now look more closely at God's role as judge in the conscience, conceived of as an internal court. Besides presiding as judge, God, Kant explains, must be thought of as imposing the moral obligations "as his commands" (MM 6:439). But this is perplexing, for the commanding, authoritative nature of the moral law is something that the moral law according to Kant is supposed to enjoy by itself without any divine support. Indeed, it is the *respect* for the moral law that grounds the recognition of the necessity of complying with our moral duties (e.g., *Groundwork of the Metaphysics of Morals* 4:400). The second characteristic of the ideal person, the internal judge, is that he "must also have all power (in heaven and on earth) in order to give effect [*Effekt*] to his laws" (MM 6:439). What is the "effect" of these laws? Kant does not explicitly explain what he has in mind here, but it seems that Kant's idea is that the judge has "power" insofar as he "pronounces the sentence of happiness or misery, as the moral results of the deed" (6:438, footnote). In other words, it seems that the power of the judge consists in the ability either to punish us, that is, to make us suffer bad conscience or guilt in case our deed is found in violation of the moral law, or to make us, or rather allow us, to rejoice "at having escaped the danger of being found punishable" (MM 6:440). This, too, is quite bizarre, for the idea seems to be that in

order to make sense of the happiness or misery that results from the judicial procedure of conscience, one *must* think of God as in some way responsible for the ensuing affective state, whether negative or positive.

The general move Kant seems to be making here is reminiscent of the argument found already in the first *Critique* (A818–819/B846–847), where, while the authority of morality is argued to be based in pure reason itself and not in God, God is nevertheless seen as required by morality as a practical *postulate* (see first *Critique*, e.g., A811/B839ff.; second *Critique* 5:122ff.; cf. *Religion* 6:5ff.) so as to render the idea of actually achieving man's highest ends not an empty hope.⁶ God is thereby not theoretically proven—where “theoretical cognition” is that “through which I cognize what exists” (first *Critique* A633/B661)—but is “led to” as a result of the immanent demands of practical reason itself. Similarly, here, in the discussion of conscience, Kant argues that while the conscience does not provide us with theoretical grounds to posit the actual existence of God, it nevertheless “points” us toward God as a result of the internal, necessary demands of conscience itself, namely, that we are required to think of God as the internal judge in order to avoid self-contradiction in the conscience.

All of this, in my view, amounts to a rather disappointing account, and as an attempt to provide a quasi-phenomenologically-based psychology of conscience, it is, I think, unsuccessful. Despite these complications, it remains that for Kant, (retrospective) bad conscience or guilt involves the pain of failing to do as the moral law demands.⁷ It might thus seem that all that is required of the internal judge is to determine the performed action's alignment or misalignment with the particular demand of morality relevant at the time of action. But this cannot be right, for this account makes

⁶ As Anat Matar puts it, as soon as God is removed from Kant's moral philosophy, he is reintroduced in order to ensure the very possibility of moral discourse: God is posited as soon as he is erased (Matar 2019, 21–22).

⁷ See also Gasdaglis 2019.

no mention of a question I have so far omitted, the question of *responsibility*: just as it is standard for actual court procedures to involve determination of the accused's degree of responsibility before reaching a verdict, so, too, must the internal court of conscience determine responsibility before it can decide upon condemnation or acquittal. Specifically, the question of whether the agent was *free* at the time of action—freedom being a necessary condition of moral responsibility—must be part of Kant's discussion of conscience.

And indeed it is, for right at the beginning of the section in the *Metaphysics* dedicated to the conscience, Kant claims that part of what takes place before the tribunal of conscience is “the internal *imputation* [*Zurechnung*] of a *deed* as a case falling under a law” (MM 6:438). And earlier in the *Metaphysics*, Kant clarifies that “*Imputation* [*Zurechnung*] (*imputatio*) in the moral sense is the *judgment* by which someone is regarded as the author (*causa libera* [free cause]) of an action, which is then called a *deed* (*factum*) and stands under laws” (6:227). In other words, imputation is Kant's concept for the assigning of moral responsibility.⁸ It thus seems that a crucial step the court of conscience takes is to determine whether the action is imputable at all, that is, whether the agent can be held morally responsible for it, which requires determining whether the agent was a “free cause” of the action, whether he or she acted freely. This indicates that to understand the working of the conscience, we must turn to look at Kant's account of freedom in the *Metaphysics*.

II. Freedom and Moral Responsibility: A Look at the *Metaphysics of Morals*

As is well known, according to Kant's transcendental idealism, events occurring in space and time—the realm of

⁸ As Saunders claims, “Kant does not often speak of responsibility” but more often of “imputability” (Saunders 2019, 139).

phenomena—are governed by a strict determinism which leaves no room for freedom. Freedom, for Kant, therefore, can only be found at the level of the noumenal, where, while it cannot be theoretically known and can never be given in experience, it can be thought without contradiction with the determinism of nature and is, moreover, required practically for the possibility of morality. This, in brief, is Kant's concept of transcendental freedom. In the course of his career, however, Kant's specific views on the nature of freedom and the will underwent various subtle and complicated changes. Here, though I will first consult other texts, I wish to examine his views on freedom and the will as they are laid out in the *Metaphysics*, for, as we saw, in this work he also provides us with the most worked-out articulation of his views on the conscience and guilt. The questions to ask now are: On what basis can the internal courtroom of conscience determine the agent's freedom? On what basis can it impute the action?

If we turn first to two of Kant's earlier texts to take a closer look at his conception of imputation and the freedom it requires, we find the following. In the first *Critique*, in the context of his discussion of the third antinomy, Kant claims:

The transcendental idea of freedom is far from constituting the whole content of the psychological concept of that name, which is for the most part empirical, but constitutes only that of the absolute spontaneity of an action, as the real ground of its imputability [*Imputabilität*]; but this idea is nevertheless the real stumbling block for philosophy, which finds insuperable difficulties in admitting this kind of unconditioned causality. Hence that in the question of freedom of the will which has always put speculative reason into such embarrassment is really only transcendental, and it concerns only whether a faculty of beginning a series of successive things or states from itself is to be assumed. (A448/B476)

In other words, here Kant puts forward a very strong claim, namely, that for an action to be morally imputable to an agent, the agent has to be considered free in the noumenal, that is, transcendental sense. By transcendental freedom Kant means our ability to determine our actions in complete independence of nature. Specifically, a transcendently free action is independent causally from everything empirical, whether internal (natural impulses or representations) or external (forces working on the body), but beginning spontaneously of itself. This freedom cannot ever be known to obtain by us, though it is thinkable in the sense that we can form a concept of it that does not contradict our conception of nature. Similar ideas are formulated later, in Kant's second *Critique*, where he claims that

transcendental freedom . . . must be thought as independence from everything empirical and so from nature generally, whether it be regarded as an object of inner sense in time only or also of outer sense in both space and time; without this freedom . . ., which alone is practical a priori, no moral law is possible and no imputation in accordance with it. (5:97)

So here again, we are told that for an action to be imputed to an agent, it must be transcendently free. It is here that Kant rejects the compatibilist position as he understands it (he calls it "psychological" freedom (5:97): being causally determined to act, even if the causes are of the "right" kind, for example, representations of what one ought to do, amounts to little more than what he refers to as "turnspit" freedom, and thus does not leave room for genuine, transcendental freedom (5:97).

Now we can better grasp the acuteness of the problem facing us: if the pronouncements of conscience require a determination of imputability, and if the latter involves transcendental freedom, how can the court of conscience possibly determine this imputability, given that "no instance corresponding to it [i.e., to transcendental freedom] can be given in any possible experience" (MM 6:221)?

No self-reflection or soul searching can provide the mind with any cognition of the action's transcendental freedom. Consequently, no imputation can ever be justified and therefore no verdict and sentence of guilt.⁹ The justified consciousness of guilt proves to be impossible. The best that can be had is a sham verdict of guilt pronounced by a sham court.

In the first *Critique*, Kant grasps the implications of this view and claims in a footnote that

The real morality of actions (their merit or guilt), even that of our own conduct, therefore remains entirely hidden from us. Our imputations [*Zurechnungen*] can be referred only to the empirical character [i.e., the person's character as manifested in specific deeds]. How much of it is to be ascribed to the pure effect of freedom [*reine Wirkung der Freiheit*],¹⁰ how much to mere nature and innocent defects of temperament or to its happy constitution (*merito fortunae*) this no one can discover, and hence no one can judge it with complete justice. (A551/B579, footnote)

Kant thus seems to raise doubts about the very ability of anyone—even the subject herself—to judge whether an action is really imputable and thus, consequently, whether guilt is justifiable or not. This skepticism, for obvious reasons, undercuts the entire notion of the conscience as an internal court, for here Kant is basically saying that whatever decision such a court arrives at, its pronouncements can never be considered just; every imputation can only concern the empirical character, that is, my character as revealed empirically, that is, the rule-governed manner in which my actions are related

⁹ Kant distinguishes degrees of imputability on the basis of the degree of difficulty the action at issue posed to the agent and the stringency of the relevant duty (MM 6:228). Thus, an action the performance of which was supererogatory and involved, in addition, overcoming significant obstacles deserves great merit. The problem I raise in the text, however, is whether any action at all can ever be imputed in the moral sense.

¹⁰ This crucial phrase given here in German is missing from the 1998 *Cambridge Edition* of the first *Critique* at my disposal.

to my past actions, taking into account the conditions under which I acted, and so on, all of which are causally determined appearances in space and time (see A549/B577). Thus, an action performed by me that goes against the categorical imperative of pure reason cannot justly be imputed to me if I did not freely choose to perform it but was rather caused to do so by empirical factors. The same holds with respect to all my past actions, on the basis of which I can be seen as callous or wretched by others or even by myself. But, as Kant clarifies in the quotation just given, as long as it is not clear whether my corrupt empirical character and the actions that spring therefrom are a *pure* effect of freedom or of nature, no action can be genuinely imputed to *me*.¹¹ Is it possible that in the *Metaphysics*, Kant is presupposing a *different* conception of imputation, one not based on transcendental freedom, so as to make sense of the possibility of guilty conscience?

To answer this question, we have to introduce Kant's concept of *practical* freedom, originally introduced in the first *Critique*. While transcendental freedom is not something we can ever have cognition of, the practical concept of freedom "proves its reality," since, as Kant puts it:

Practical freedom can be proved through experience. For it is not merely that which stimulates the senses, i.e., immediately affects them, that determines human choice, but we have a capacity to overcome impressions on our sensory faculty of desire by representations of that which is useful or injurious even in a more remote way; but these considerations about that which in regard to our whole condition is desirable, i.e., good and useful, depend on reason. Hence this also yields laws that are imperatives, i.e.,

¹¹ It is worth noting that even when an agent's action conforms to the moral law, there is no telling according to Kant whether they deserve genuine moral merit; it is possible they were caused not by genuine moral considerations but by self-love (see *Groundwork of the Metaphysics of Morals* 4:407).

objective laws of freedom, and that say what ought to happen, even though perhaps it never does happen. (A802/B830)

So practical freedom *is* something of which we have experience: we have experience of instances where we struggle against our desires (nature), overcome them, and manage to do what we are required to do by the laws of pure reason, that is, to do what is *morally* required of us. This shows that even though, like that of animals, our choice can be affected by our natural impulses, in contrast to that of animals, our choice is not simply determined by these but can be determined by the will, that is, by the laws of pure practical reason.¹² And since Kant's discussion in the *Metaphysics* seems to proceed on the basis of the practical concept of freedom and seems to put the transcendental one aside, it arguably makes sense to understand his talk of freedom in his definition of imputation in the *Metaphysics* as referring to the *practical* concept of freedom.

Thus, in the *Metaphysics*, Kant begins his discussion in the section "Preliminary Concepts of the Metaphysics of Morals" by distinguishing the "transcendent" freedom of which "we cannot obtain any theoretical cognition" from

reason's *practical* use of the concept of freedom [which] proves its reality by practical principles, which are laws of a causality of pure reason for determining [*bestimmen*] choice [*Willkür*] independently of any empirical conditions (of sensibility generally), and prove a pure will [*Willen*] in us, in which moral concepts have their source. On this concept of freedom, which is positive (from

¹² Importantly, this experience of practical freedom does not prove the positive conception of intelligible or transcendental freedom: as Kant clarifies in A803/B831, it is possible that the reason which issues its moral commands and to which we listen is in turn caused by some other factors, unknown to us. This is why, even if we experience the triumph of reason over the senses and thus witness our practical freedom, this does not mean that our reason was indeed free in the positive, transcendental sense, that is, utterly uncaused (fully spontaneous).

a practical point of view), are based unconditional practical laws, which are called *moral*. (MM 6:221; first emphasis added)

In line with this introduction of the practical concept of freedom, Kant goes on later to define more explicitly the central relevant concepts. He thus distinguishes the will (*Wille*) from the power of choice, or simply choice (*Willkür*),¹³ and explains, rather surprisingly, that the will cannot be called free or unfree, for the will is “practical reason itself” and is thus concerned immediately with “giving laws for the maxims of actions,” that is, for the practical principles on which we can act and which are adopted by choice, or “proceed . . . from choice” (6:226). In contrast, (human) choice is free: “Only *choice* can therefore be called *free*” (6:226). Why? Kant here distinguishes a negative from a positive conception of freedom and claims that in contrast to that of animals, human choice “is a choice that can indeed be affected [*afficirt*] but not determined [*bestimmt*] by impulses. . . . Freedom of choice is this independence from being determined by sensible impulses; this is the *negative* concept of freedom” (6:213; emphasis added; cf. A533–534/B561–562). On the other hand, the *positive* concept of freedom is that “of the ability of pure reason to be of itself practical” (MM 6:214), that is, of the ability of pure reason or the will to determine one’s actions by determining the power of choice to incorporate the moral law into one’s maxims. As Kant writes, “That choice which can be determined by *pure reason* is called free choice” (6:213). Notice that both the positive and the negative concepts of freedom introduced here are concepts of *practical* freedom, for they do not involve spontaneous freedom from all causality.

It thus might seem that since we, qua human beings, possess free choice in this practical sense, as long as the individual “was

¹³ As is well known by Kant scholars, this distinction, though inexistent at the time of the first *Critique*, came to be employed later in his career, especially in his *Religion*, to which I shall turn later.

in his senses, that is, had the use of his freedom" (5:98)—which means that the individual was not inebriated, was not held at gun-point, and so on—an individual's conscience should impute the performed deed to the individual and, in case an immoral deed is at issue and the "court proceedings" thus indicate, determine him guilty.

There is, however, a problem here: if to be free in the practical sense means to be able to overcome one's desires and do what's right, then to be affected by one's impulses in such a way that one fails to overcome them and as a consequence performs an immoral deed would seem to indicate *lack* of ability to overcome them, hence lack of practical freedom, at least at the time of action, which would then undermine imputation and all guilt. It thus seems that whether imputability is conditioned by transcendental or practical freedom, the upshot is the same: guilt is unjustified.

One possible way to think about immoral yet imputable action is to think about an agent who is not determined by the natural impulses to act as they tempt him to. The agent, in this predicament, stands back, so to speak, from both the moral incentive and the natural impulses and decides with which one to side. If he chooses to act immorally, we can then say that he nevertheless had the power to do the right thing—it was up to him, and yet he chose to act immorally. His action is therefore imputable to him.

The first problem with this suggestion is that it would give rise to the following question: if it is possible for an agent to choose *not* to act on a sensuous impulse but instead do what is right, why doesn't he? How could imputable immoral action be explained? If neither morality nor the impulses are strong enough to determine one's choice, why does the agent choose to act immorally? Second, this way of thinking about the agent's predicament would involve attributing to the agent's power of choice a liberty of indifference of the sort which Kant denies: "freedom of choice cannot be defined . . . as the ability to make a choice for or against the law (*libertas indifferentiae*), even though choice as a *phenomenon*

provides frequent examples of this in experience” (MM 6:226).¹⁴ That is, even though it might appear from the outside, so to speak, that we are free to choose to act morally or not as we please, this is not the case. Rather, “Only freedom in relation to the internal lawgiving of reason is really an ability; the possibility of deviating from it is an inability” (6:227). Choice is free because we have the ability to determine it with the commands of reason, not because we can also choose to act immorally. When we act immorally, it is a result of a *failure* to exercise this ability. But how is this failure or “inability” to be made sense of in a manner that preserves its imputability?

To try to make sense of this, let us first clarify the meaning of “determine” in the *Metaphysics*’ explanation of the positive and negative concepts of practical freedom just introduced. A very plausible suggestion is that by “determine” Kant means being sufficient to produce the consequence or effect (Engstrom 2002, 295). If this is the case, then when Kant says that choice cannot be determined but only affected by the sensible impulses, he means that by themselves, those impulses cannot make the power of choice choose and thus bring about an action. A first possible explanation for why this is so—for why the impulses cannot determine choice—is that following what Henry Allison has called the incorporation thesis, “an inclination or desire does not *of itself* constitute a reason for acting” (Allison 1990, 40) but becomes one only when incorporated into one’s principle of action, that is, into a maxim. It then becomes possible to say that according to the negative concept of freedom of choice, the natural impulses by themselves are not sufficient to bring about an action but must first be taken up by one’s power of choice and incorporated into one’s maxim; only then do they become a reason for action and able to bring it about. This

¹⁴ Kant, it seems, is contradicting himself here, for earlier in the text, he claims that the power of choice, as belonging to the faculty of desire, involves the ability “to *do or to refrain from doing as one pleases*” (6:213).

explanation, however, cannot be satisfactory, for according to the *positive* concept of freedom, the will *can* determine choice, that is, can be sufficient to bring action about, and yet here, too, the moral law would first have to be taken up by choice; it does not just force itself upon the agent but must first be incorporated into the agent's maxims.¹⁵

So what could it mean to say that the natural impulses cannot by themselves bring about action? In order to look at this problem more closely, let us turn briefly to Kant's concept of moral *strength*, which is how he understands *virtue* (MM 6:405). Virtue, says Kant, is "the strength of a human being's maxims in fulfilling his duty,"¹⁶ where "strength of any kind can be recognized only by the obstacles it can overcome"; obstacles which, in the case of virtue, are man's "natural inclinations, which can come into conflict with the human being's moral resolution" (6:394). How is the concept of strength to explain the possibility of an immoral yet imputable action? In response to this problem, Stephen Engstrom argues that "though sensible impulses are not causes that are by themselves sufficient to determine this power [of choice], it might still be possible for them, under certain conditions, to have an influence upon its exercise" (Engstrom 2002, 296). What are these "certain conditions"? According to Engstrom, these refer to "the presence of *weakness* in the subject's will (practical reason) sufficient to leave the power of choice liable to be influenced by the sensible impulse" (299; emphasis added). Allen Wood, on whom Engstrom presumably relies,

¹⁵ It is possible that a slight change of mind has occurred between the time of *Religion* and that of the *Metaphysics*, for while in the latter text, as we saw, Kant holds that the power of choice *can* be determined (at least by reason), in the former text, Kant claims that "the power of choice [*Willkür*] has the characteristic . . . that it cannot be determined [*bestimmt*] to action through *any* incentive [*Triebfeder*] *except so far as the human being has incorporated it into his maxim* . . . only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of freedom). But the moral law is itself an incentive." (*Religion*, 6:24; first emphasis added).

¹⁶ The formulation Kant uses here is admittedly odd insofar as it attributes strength not to the person or his or her spirit but to the person's maxims or principles. The idea, nevertheless, I think, is clear enough: to possess strong maxims just is to be strong oneself in one's ability to act according to the moral law.

makes a similar argument and claims that to be free is to have the power to determine oneself according to the moral law. But this does not mean that one invariably will be able to *exercise* this power. When, indeed, one fails to exercise one's power, this "is not due to a lack of freedom"—the failure, rather, should be regarded as due "to a certain weakness or lack of power, leading to a lapse in exercising our freedom" (Wood 1984, 81). This makes it possible to retain responsibility, as Engstrom puts it: "So the choice of action contrary to the moral law does not by itself imply that in that choice the subject lacks the capacity to choose to act according to the law, nor therefore that the subject is not free and accountable for the action" (Engstrom 2002, 296).

If we wish to retain Allison's incorporation thesis, we can combine it with Kant's claims about determination in the following way: the will can determine the power of choice in the sense that when the will is strong, when the person, that is, is morally virtuous, the presence of the moral incentive is sufficient to make choice incorporate the moral law into its maxim, leading to moral action; conversely, when the will is weak, the presence of natural impulses is sufficient to make the power of choice incorporate heteronomous inclinations into its maxim, leading to immoral¹⁷ action.¹⁸

It seems it is thus possible to eat one's cake and have it, too, that is, to explain the performance of immoral action while maintaining moral responsibility: one is responsible for the immoral action one performed and can rightly charge oneself with it because one was free at the time of the action, that is, one had the power to act out of respect for the moral law, and yet one acted immorally because one

¹⁷ It should be noted that moral weakness as explained here does not *necessarily* have to generate immoral action; it is possible for action motivated by natural impulses to be in "legalistic" conformity with the moral law.

¹⁸ An uneasiness that may arise from these formulations—one that can be traced back to Kant's own notion of "determination"—is that they involve thinking quasi-causally about the incorporation of incentives, where this, for Kant, is supposed to transpire outside the realm of experience and thus outside the relation of cause and effect. I think a possible way to address this is to appeal to a notion of causality that is independent of the conditions of experience such as place and time—a form of intelligible causality.

failed, out of weakness, to exercise this power, because when one is weak, the natural impulses *can* determine the power of choice to incorporate them into its maxims, thus bringing about immoral action.

One unsatisfactory element of this appeal to weakness is that it has the implication that in contrast to Kant's assertion (MM 6:213), reason cannot *simply* be said to be able to determine one's power of choice, either, for it requires the additional condition of moral strength. But even if we put this (minor) problem aside, the suggested solution cannot work. For consider the following. Earlier, we asked for a clarification of the possibility of imputable immoral action. The natural impulses by themselves cannot determine choice, and we were told that the missing ingredient needed to explain their possible power over choice is weakness: in the presence of weakness, they can influence choice and lead to immoral action (or merely action lacking any moral worth one way or the other). But this in effect means that in the presence of weakness, natural impulses can *determine* choice. For if we were to deny this, then we would again be faced with the task of making sense of the possibility of imputable immoral action and be forced to invoke a further ingredient that in cooperation with the impulses *and* weakness determines choice of immoral action, and no such further ingredient, it seems, is on offer. It thus seems that immoral action is determined, and thus made sense of, by the working of the natural impulses and weakness. But if this is the case, if the agent was determined to act as he did when acting immorally, on what grounds can we hold him responsible for their action?

Kant, perhaps, would insist that the agent was nevertheless free at the time of action and could thus justifiably be held responsible, because the agent had *the power* to act autonomously: he "could have acted instead on a priori [i.e., moral] motives" (Wood 1984, 80). But what is meant by "could have" here? In what sense could the agent have acted on moral motives? And if he could, why didn't he? The problem is that at the time, the agent, it seems, could *not*

possibly have exercised this power to act autonomously given the way his choice was affected by the natural impulses and weakness. But, importantly, it would be unfair to hold a person responsible for not doing something that he *could not have done* at the time. In his attempt to defend Kant on this score, Wood gives the following analogy:

[A] swimmer may have the power to save himself if he falls into deep water. But he has no power, only a possibility, of drowning in the same eventuality. Although he has the power to swim, he may drown if he does not exercise that power effectively, due (say) to confusion or panic. In the latter case, his possibility of drowning is due to a kind of weakness or lack of power, though not to a lack of the power to swim. (Wood 1984, 82)

The problem, however, is that it would be utterly absurd to say of a drowning swimmer that instead of drowning, he *could have* used his power to swim to save himself—and is therefore responsible for his drowning—but nevertheless didn't because of "a kind of weakness or lack of power." And it is precisely because of this weakness that it would be wrong of us to hold such a swimmer responsible for his failure to swim back to the top on this specific occasion; after all, the swimmer in the example did not freely *choose* to be confused or to panic, and under the condition of panic, his "power to swim" was rendered totally ineffective.

And yet—Kant holds—guilt persists. As he puts it in the second *Critique* (I quote at some length):

The judicial sentences of that wonderful capacity in us which we call conscience are in perfect agreement with this [i.e., with holding oneself responsible despite the empirical necessity of the action]. A human being may use what art he will to paint some unlawful conduct he remembers as an unintentional fault . . . as something in which he was carried away by the stream of natural

necessity—and to declare himself innocent of it; he nevertheless finds that the advocate who speaks in his favor can by no means reduce to silence the prosecutor within him, if only he is aware that at the time he did this wrong he was in his senses, that is, had the use of his freedom; and while he *explains* his misconduct by certain bad habits, which by gradual neglect of attention he has allowed to grow in him to such a degree that he can regard his misconduct as their natural consequence, yet this cannot protect him from the reproach and censure he casts upon himself. This is also the ground of repentance for a deed long past at every recollection of it, a painful feeling aroused by the moral disposition. (5:98)

And a bit later in the text, Kant explains why this is so, why judgments of guilt, despite determinism, remain “founded”:

This could not happen if we did not suppose that whatever arises from one’s choice [*Willkür*] (as every action intentionally performed undoubtedly does) has as its basis a free causality, which from early youth expresses its character in its appearances (actions); these actions, on account of the uniformity of conduct, make knowable a natural connection that does not, however, make the vicious constitution of the will [*die arge Beschaffenheit des Willens*] necessary but is instead the consequence of the evil and unchangeable principles freely adopted [*die Folge der freiwillig angenommenen bösen und unwandelbaren Grundsätze*], which make it only more culpable and deserving of punishment. (5:100)

One way to read this paragraph, which, as we shall see, chimes considerably with ideas Kant develops in his later *Religion within the Boundaries of Mere Reason*, is that even though every specific misdeed can be explained in a way that seems to expunge all culpability, the immoral agent would still feel guilty, *and rightly*

so, because an agent's wrong actions are nevertheless ultimately grounded in freedom—the freedom that adopts evil principles and thus constitutes one's will as vicious.¹⁹ In other words, culpability is to be found not on the level of discrete actions but on a much deeper level where one freely adopts one's basic and “unchangeable principles,” one's *fundamental* maxims.

One thing that we have to notice here is that if indeed moral fault lies not with the performance of any specific misdeed but with a more fundamental fault, then it would seem to follow that *this* is the immorality that guilt tracks rather than the specific misdeed itself. But we should underscore the peculiarity of this result, for guilt is typically experienced by us in light of, and moreover concerns, a *specific* empirical commission or omission. Thus, for example, the guilt I might feel as a result of not divulging the entire truth to my friend, it seems, would not be about my failure to develop sufficient moral fortitude but about my specific moral failure. As we shall see in the next chapters, both Schelling and Schopenhauer hold similar views where guilt for specific misdeeds has the specific misdeed only as its apparent object: the real object of one's guilt lies at a deeper metaphysical level. But to elaborate on Kant's thinking here, we have to turn to *Religion*.

III. “Innate” Guilt and Freedom in *Religion within the Boundaries of Mere Reason*

In this section, I will argue that it is in the earlier *Religion within the Boundaries of Mere Reason* (first published as a whole in 1793)²⁰ that we can find an answer to our question regarding the

¹⁹ I am thus sympathetic to Alenka Zupančič's analysis: “The guilt that is at issue here is not the guilt that we experience because of something we may or may not have done . . . we might say that guilt is the way in which the subject originally participates in freedom” (Zupančič 2000, 26–27).

²⁰ Hereafter cited as *Rel*. All references to *Religion within the Boundaries of Mere Reason* are to the standard pagination.

imputability of guilt. It is here that Kant shows how a different, deeper kind of guilt—which I call *ontological guilt*—gives reason to think we are transcendently free, and it is this freedom that can ultimately justify guilt and the experience of guilt for our specific empirical misdeeds (factual or empirical guilt).

Kant's *Religion* is not typically read as particularly concerned with, or as providing an argument for, freedom.²¹ Nevertheless, while it is indeed not Kant's stated aim to argue for freedom in this work, I argue that Kant in effect does make such an argument on the basis of our evil or corrupt nature or (what I call) ontological guilt. As should be obvious, I don't intend to provide here a thorough interpretation of the work or defend any principled approach to its interpretation. This lies beyond the boundaries of my investigation.

Broadly speaking, in *Religion*, Kant attempts to determine to what extent and which of the basic tenets of Christian orthodoxy can be seen to be supported by purely rational considerations and thus fall within what Kant calls the "pure *rational system of religion*" (*Rel.* 6:12). One of these tenets is the Christian idea of original sin (6:31), which Kant refers to as "radical evil in human nature" (6:19)—an evil that is not contingent but rather "applies to [man] considered in his species" (6:32). In other words, what Kant is referring to here is the idea of ontological guilt: a guilt that we carry qua human beings insofar as we possess a "corrupt propensity [*Hang*]" (6:32) to evil, a propensity which lies "in human nature" (6:37). Though Kant's use of the concept of propensity is confusing,²² he makes it clear that he does indeed think that such

²¹ Though see Guyer 2009, which I will discuss in a later footnote.

²² Thus, the concept of propensity seems to imply a natural, and thus unchosen, disposition to or for something, yet Kant claims that a propensity can also be thought of as acquired (*Rel.* 6:29); second, while the propensity to evil involves the propensity to adopt as one's maxim the subordination of the moral law to self-love, Kant seems to hold that the propensity just *is* this fundamental maxim, as when he writes that "in this propensity the maxim has been assumed to be evil" (6:31–32); third, though the propensity is supposed to have been adopted by the human being by means of a free, intelligible deed, Kant writes that the propensity *itself* is the deed ("The propensity to evil is a deed" (6:31). I will therefore assume that the propensity to evil is the subverted fundamental maxim which the human being is responsible for freely adopting. Furthermore, I agree

a propensity is real and adds, on the basis of anthropological data he adduces (6:33), that “according to the cognition we have of the human being through experience, he cannot be judged otherwise” (6:32).²³ This propensity of “corruption of the human heart” (6:30) or “depravity” (6:30) is the third and most wicked propensity to evil in man, and it is this on which I will focus.²⁴

What this propensity to radical evil consists in (*Rel.* 6:36–37) is the subordination of the moral law to the principle of self-love as one’s fundamental maxim of action. Rather than first asking herself whether a course of action is morally required or permissible and then following suit without concern for her own personal interests, an individual in whom this propensity to evil has taken root has made it her general, fundamental practical policy to be such that she should first determine whether the action under consideration aligns with or at least does not harm her own self-interests, and only if the answer is positive should she then turn to examine whether it is also morally permissible or required. Practical precedence is thus given to one’s self-interests. As Kant clarifies, this evil does not imply that one’s actions are necessarily immoral in the sense of contradictory to the moral law, only that they are not done out of respect for the moral law and are at best merely in conformity with it. In the latter case, “the empirical character,” that is,

with Bernstein that “Most of the time [Kant] writes as if there is no significant difference between a disposition (*Gesinnung*) and a propensity (*Hang*)” (Bernstein 2002, 26), and I, too, use the two terms interchangeably in my discussion—clearly distinguishing them will not affect the argument I attribute to Kant. I nevertheless also agree with Bernstein’s claim that the two must be different insofar as Kant holds that the disposition can be either good or evil while the propensity in Kant is *only* to evil (Bernstein 2002, 26). Though I will not go into detail, I think that to freely choose one’s propensity to evil is to acquire an evil disposition. To acquire a good disposition—that is, to adopt the moral law as one’s fundamental maxim—is something Kant thinks is possible, albeit not “through human forces” (*Rel.* 6:37).

²³ Kant’s attempt to ground his judgment of the species as such has been addressed by several commentators (see, e.g., Wood 1970; Allison 1990; Pasternack 2014). I will not go into this issue here and in the rest of my discussion assume for argument’s sake that Kant’s judgment is plausible. However, I return to this question briefly in the chapter’s conclusion.

²⁴ The other more benign forms of the propensity to evil are explained in *Rel.* 6:29–30.

the person's character as manifested in specific deeds, "is then good but the intelligible character"—governed by the fundamental principle of choice—is "still evil" (6:37).²⁵

Importantly, by calling it "natural" or "innate," Kant means to say that this propensity "belongs to the human being universally" (*Rel.* 6:29), but not in the sense that the *nature* of the human being is to blame for it, for this would "stand in direct contradiction to the predicates *morally* good or *morally* evil if [nature is] taken to mean (as it usually does) the opposite of the ground of actions [arising] from *freedom*" (6:21). Rather, this propensity to evil "must be capable of being imputed to the subject as itself guilty of it [*als selbst verschuldet ihm muß zugerechnet werden können*]" (6:35).²⁶ As we know, to say an action or a condition is imputable to the agent means that the agent is responsible for it in a sense that can justify blame and condemnation. In other words, the subject must be seen as morally *responsible* for this propensity to evil which constitutes the subject's innate guilt. Indeed, Kant holds that this "*innate* guilt [*angeborene Schuld*]" is "*deliberate* guilt [*vorsätzliche Schuld*]" (6:38).²⁷ Thus, in order for the human being to be guilty by

²⁵ I will get back to the notion of intelligible character toward the end of the chapter, but for now, the idea that it should be closely related to, if not identified with, one's most fundamental maxims should suffice to make sense of my argument.

²⁶ Bernstein rightly complains that this is rather bizarre; it is not clear how something can both be a propensity and yet be freely willed by us (Bernstein 2002, 30–31). This, I take it, is also the force of David Sussman's complaint in his review of Pablo Muchnik's analysis of Kant's *Religion* that "Despite Muchnik's efforts, Kant's position seems to be trapped in a circle. As an aspect of human nature, radical evil has to assume the role of something like a psychological tendency that we are always confronting in ourselves. Such a tendency may explain why we make immoral choices, but only at the price of their freedom and imputability. On the other hand, if we are to sustain imputability, radical evil must take the guise of an activity we always find ourselves freely engaging in. But then the appeal to radical evil explains nothing about our basic misuse of freedom, since such evil simply is that misuse itself" (Sussman 2010). I believe Kant's view is that the propensity is universal in man, as is its free adoption as a fundamental maxim, and it is this latter act for which we are morally responsible. What explains the universality of this latter deed is not, however, satisfactorily addressed by Kant in my view, and Bernstein plausibly contends that it in principle *cannot* be satisfactorily addressed (Bernstein 2002, 35; though see Pasternack 2014, chap. 3). These difficulties precisely push the later Schelling and Schopenhauer to think of the agent as a *causa sui*.

²⁷ Kant claims that our guilt is deliberate—we intentionally adopted the evil maxim. But—and here a question arises—how can adopting the evil maxim be an immoral act

“nature,” the human being must be considered morally responsible for the fundamental maxim, for this original sin, and it cannot “lie in any object *determining* the power of choice [*Willkür*] through inclination, and not in any natural impulses, but only in a rule that the power of choice itself produces for the exercise of its freedom” (6:21). In other words, the human being is considered by Kant to be guilty—morally responsible—for his or her fundamental or ontological guilt, and this responsibility can only be made sense of with the assumption of *freedom*.

And here we seem finally to have an answer to our question about the justifiability of factual or empirical guilt: our particular actions lack moral worth at best or are downright wrong, and we are guilty for them *because we are responsible for our basic “corrupt nature,” for our ontological guilt*. And what is it that makes our responsibility for this ontological guilt possible? We are responsible for this fundamental or ontological guilt by virtue of our *free* choice of our evil fundamental maxim; by virtue of an “intelligible deed” (6:31). As Kant puts it, “The disposition [*Gesinnung*], i.e. the first subjective ground of the adoption of the maxims . . . must be adopted through the free power of choice, for otherwise it could not be imputed [*Sie selbst aber muß auch durch freie Willkür angenommen worden sein, denn sonst könnte sie nicht zugerechnet werden*]” (6:25). In other words, we freely choose to determine ourselves so as to be governed by this ultimate practical policy, for otherwise we could be held morally responsible neither for our “nature” nor—I add—for the resultant empirical actions which are grounded in it, whether immoral or amoral.

So on my reading, Kant’s thinking, in effect, amounts to the following argument: since we are evil by nature and “cannot be judged otherwise” (*Rel.* 6:32), and since, moreover, Kant thinks, this evil is

unless we *know* that it is evil to do so? There must be some recognition of the fact that this maxim is evil and so is its adoption in order for the original or ontological guilt to be genuinely attributable to us, for without knowledge, there is no guilt. Kant, as far as I know, says nothing about this.

“moral,” that is, imputable, something “a human being can be held accountable for” (6:32), it follows that we acquired this evil through a free exercise of our power of choice insofar as we freely chose to adopt as our fundamental maxim one where self-love takes precedence over morality. Thus, the moral evil in us—our ontological guilt—gives reason to believe that each of us has actually and freely exercised his or her power of choice in this way. The argument is essentially of a transcendental sort: a necessary condition for our being morally evil is our having freely performed the intelligible deed Kant mentions. Ontological guilt implies transcendental freedom.²⁸

How do factual guilt and ontological guilt stand with respect to each other in light of Kant’s discussion in *Religion*? First, we can be factually morally guilty for our deeds only because we are morally responsible for our original sin, for our ontological guilt; it is in the free choice of our original maxim that our true responsibility lies. Had we not been morally responsible for our original guilt, we would not have been morally responsible for any of our actions, grounded as they are in our fundamental maxim. Put differently, the imputability of our actions is ultimately grounded in the imputability of our ontological guilt. Second, we would not

²⁸ Paul Guyer presents Kant’s *Religion* with the following question: “how can Kant suddenly [i.e., in *Religion*] maintain that we are free to choose either good or evil, and always free to choose the other no matter what we have, as it appears phenomenally, previously chosen, when he has previously argued [in the *Groundwork* and the second *Critique*] that the moral law is the causal law of the noumenal self?” (Guyer 2009, 200). Guyer’s response is that in *Religion*, Kant “appeals exclusively to the principle that *ought implies can*” (200), as well as the corollary principle that *ought* does not imply *does* (201), to argue that since we recognize the validity of the moral law, we see that we can correct our ways (or relapse) by adopting the moral law as our fundamental maxim (or abandon it in favor of self-love) (see, e.g., *Rel.* 6:45). It thus seems that freedom in *Religion* can be established not on the basis of our ontological guilt but on the basis of the “*ought implies can*” principle. But I don’t think this principle is robust enough to support the conclusion that we can act freely in a transcendental sense. And indeed, Kant in *Religion* does not think that we actually *can* correct our ways—at least not on our own, but we are in need of divine support or “supernatural cooperation” (6:44): “what we can do is of itself insufficient” (6:45). Furthermore, Guyer’s interpretation does not show that we have indeed acted freely but only that we can. On my reading, the *Religion* argument is that each of us has in a sense already performed a free deed.

have acted in a way that lacks genuine moral worth or is flatly immoral to begin with had it not been for our fundamental choice of maxim; our ontological guilt thus serves to *explain* our behavior.²⁹

Indeed, according to *Religion*, what explains our immoral behavior is not weakness but the evil *underlying* weakness. In discussing Stoic thought in his *Religion*, Kant claims that the Stoics sought to explain moral vice as a result of an

omission to combat [one's inclinations] since he [the Stoic] did not assume any special positive principle (evil in itself) [*an sich böses*]; since this omission is, however, itself contrary to duty (a transgression) and not just a natural error, and its cause cannot in turn be sought (without arguing in a circle) in the inclinations but, on the contrary, only in that which determines the power of choice as free power of choice (in the first and inmost ground of the maxims which are in agreement with the inclinations), we can well understand how philosophers . . . could mistake the real opponent of goodness with whom they believed they had to stand in combat. (*Rel.* 6:59)

Kant in effect claims here that to account for the dominance of one's inclinations requires realizing that it is not mere weakness, not a mere absence of the capacity to exercise a power that is to be invoked as the most fundamental explanation for immoral deeds, but a "special positive principle (evil in itself)." Indeed, in a footnote to his discussion of the Stoics, Kant explicitly *criticizes* the idea that "the presence of moral evil can be very easily explained, namely, by the power of the incentives of sensibility, on the one

²⁹ This latter idea, however, runs into the familiar problem of explaining how the intelligible or noumenal realm can be understood to be affecting the empirical one. Bernstein prefers to think of the disposition as "informing" rather than "causing" the specific maxims we adopt and the actions we perform in specific situations (Bernstein 2002, 25), but I don't think this is of real help, for the question of how an intelligible realm can "inform" the phenomenal realm is not significantly less mysterious than if we think of the relation in terms of causation.

hand, and the impotence of the incentive of reason (respect for the moral law) on the other, i.e. by *weakness*" (*Rel.* 6:59, footnote). It is *evil* which is responsible for corrupting one's maxims in the first place, for the agent's free choice of maxims which "are in agreement with the inclinations," and it is because we are "in secret agreement" (6:60) with our inclinations and thus let them, instead of the moral law, govern our most fundamental maxims that we fail to attain moral strength in the first place. It thus seems that in *Religion*, Kant goes deeper in his attempt to explain immorality: weakness is not enough, but one must try to fathom the innermost reason for this weakness, and this, as it turns out, is the *evil* which is a deeper affliction and corrupts one's fundamental maxim itself.

Now, what is interesting about this original free deed of our power of choice in virtue of which we acquire our evil is that insofar as it is an "intelligible deed [*Tat*]," it is "cognizable through reason alone apart from any temporal condition" (*Rel.* 6:31): for it to be a genuinely free act, it cannot be experienced or thought of as performed in time, given that time is the formal condition for the experience of objects which fall under the category of causality and are thus subject to a determinism which leaves no room for any (non-compatibilist) possibility of freedom and responsibility. The intelligible deed must then be thought of as a *timeless deed*. Moreover, since it is not given in experience, the freedom at issue in this intelligible deed cannot be of the practical kind. It must therefore be a kind of transcendently free deed. The question then arises of how to make sense of an act that is performed "not in time" (6:39). Are we to think here of an agent who operates in a purely rational fashion in some intelligible, noumenal world, a world distinct from the empirical world of appearances?

In an attempt to deflate such a metaphysical reading, Lawrence Pasternack argues that "Kant does not trace the cause of our transgressions back to something outside of time, but rather [quoting Kant] 'all the way back to the time when the use of reason had not yet developed' [*Rel.* 6:43]" (Pasternack 2014, 116), and

adds that for Kant, we have been evil in this way “*always, from [our] youth on*” (*Rel.* 6:25; emphasis in original). Pasternack’s gloss on this is that for Kant,

some choices are assigned to us as a matter of rational deduction rather than introspection. Just as we may retrospectively ascribe to ourselves a particular maxim after we have observed how we have behaved and seek to explain it, so likewise the choice(s) that ground our moral status are “representation[s] of reason” (*Rel.* 6:39). (Pasternack 2014, 115)

But this is of no help, for when we look back at our actions and explain them by retrospectively attributing to ourselves a choice we made, we clearly do not have in mind a choice that is “not in time,” as per Kant’s formulation. Further, Kant’s mention of a time “when the use of reason had not yet developed” is made in the context of a discussion of how it might be possible to think of our fundamental choice as made *in time*, that is, *from the psychological or empirical point of view*. As Kant puts it, “if we wish to engage in an explanation of evil with respect to its *beginning in time*, we must trace the causes of every deliberate transgression . . . all the way back to the time when the use of reason had not yet developed” (*Rel.* 6:42–43; emphasis in original). A few lines later, however, Kant emphasizes that

We must not however seek an origin in time of a moral character of which we are to be held accountable, however unavoidable this might be if we want to *explain* the contingent existence of this character (hence the Scriptures, in accordance with this weakness of ours, have perhaps so portrayed its origin in time). (6:43)

In other words, Kant clearly thinks that insofar as our radical evil is imputable to us, its origin *cannot* be sought in time. The attempt to trace it back to a moment in time is a psychological temptation,

one we should resist given that such an approach would completely dissolve the moral imputability of our evil.

It is true, as Pasternack points out, that in clarifying the disposition's innateness, Kant explains that this "does not mean . . . that the disposition has not been earned by the human being who harbors it, i.e. that he is not its author, but means rather that it has not been earned in time (that he has been the one way or the other *always, from his youth on*)" (*Rel.* 6:25). But, importantly and in contrast to Pasternack, the point of the parenthesis is to clarify not the meaning of "has not been earned in time" but that of the locution "earned in time." To attribute to Kant the view that the disposition was not earned in time where that means that it was earned sometime during youth or before youth is to attribute to him a blatant contradiction, given that youth and pre-youth refer to periods of time in a person's life.

Allison, too, seeks to avoid the thicker metaphysical implications of Kant's talk of a choice not made in time so as to save Kant from a "paradoxical, if not totally incoherent, doctrine of a timeless act of self-constitution" (Allison 1990, 137). He first remarks—and as we saw, Pasternack makes an essentially similar claim—that this basic maxim that constitutes radical evil is of a "very peculiar sort," since "One does not, after all, one day deliberately resolve to 'make it one's maxim' to allow moral holidays under certain conditions. It is rather that one finds that this is how one has been behaving all along" (Allison 1990, 153). But, *pace* Allison, this should not be surprising, for if indeed the disposition "has not been earned in time," or, more minimally, cannot be thought of as having been earned in time, then it most certainly cannot be thought of as something that was earned "one day." In keeping with his "two-aspect" interpretation of Kant's transcendental idealism, Allison goes on to explain that the "act is timeless or intelligible, not in the sense that it must be regarded as occurring in some timeless noumenal world but rather in the sense that it is not to be viewed as performed at a specific point in one's moral development" (153). This, I believe,

will not do, for this is just to read “timeless” as “not performed at a specific point,” without any apparent textual justification. Indeed, Kant’s claim is absolutely general and applies to all time as such, not just to a “specific point in one’s moral development,” since, again, to ascribe any temporal qualification at all to the act would amount to undermining its imputability. This means that the act could not be thought of as having been performed at a time *preceding* the beginning of one’s moral development, either. Allison continues and explains that the adoption of the maxim, since it underlies all of one’s specific deliberations with respect to how to act, “must be presupposed by, rather than revealed, in moral reflection” (153). While correct in my view, this does not touch the problem of how to understand the non-temporal or intelligible nature of the deed of the maxim’s adoption.

Allison’s more fundamental point with respect to the alleged timeless deed is that Kant “is not really propounding a noumenalistic metaphysics but simply describing how subjects, qua responsible moral agents, must take themselves” (153). This is so because rational, responsible agents do not just adopt maxims with no rhyme or reason but find justifying reasons for their incorporation or rejection in their fundamental maxim. Furthermore, such fundamental maxim, which “can itself be the object of moral evaluation,” “provides a means for thinking about the moral life of the person as a whole” (153). The idea seems to be that *if* we are to think of ourselves as rational, responsible agents, then we must *think* of ourselves as possessing a fundamental maxim for which we are to think of ourselves as responsible. In other words, Allison’s suggestion is that thinking of ourselves as rational, responsible agents commits us *not* to the view that we in fact performed some timeless deed by means of which we freely adopted our disposition but merely to *thinking* of ourselves *as if* we are responsible for this, our fundamental maxim.

This is an attractive interpretation, but it suffers from some problems. First, it seems to be troublingly textually unsupported.

If it was indeed Kant's intention merely to describe the manner in which we must think of ourselves when we take ourselves to be responsible and rational agents, then, presumably, he would have said so. In contrast, his claims in *Religion* are made without such qualification. Kant develops his ideas on radical evil not as grounded in some necessary conditions for agency that we must ascribe to ourselves if we are to regard ourselves in a certain way but on the basis of his *cognition*, disputable as it might be, that man *is* evil and that since this evil must be imputable, man must be considered responsible for his evil disposition. As we saw him put this idea: "according to the cognition we have of the human being through experience, he cannot be judged otherwise" (*Rel.* 6:32). This is not the language one employs to express a mere "practical regulative idea" (Allison 1990, 140). Moreover, regulative talk would not help us address the question of the justification of factual guilt; after all, guilt for specific misdeeds cannot be genuinely justified if I just regard myself *as if* I am responsible for my evil nature. For it to be justified, I must really *be* responsible. Guilt, to be genuine, must at some point be grounded in reality; otherwise, it would be just a mere spinning in an "as if" void.³⁰

How, then, can we make sense of this "intelligible act" of fundamental maxim adoption? For Kant, at the end of the day, the "rational origin" of our propensity to evil "remains inexplicable" (*Rel.* 6:43; see also 6:45). But the reason for this is not that the very idea of an intelligible act is too problematically metaphysical for him. The reason, rather, is that since this propensity "must itself be imputed to us, this supreme ground of all maxims must in turn require the adoption of an evil maxim. Evil can have originated only from evil" (6:43). "Yet the original disposition," Kant holds, "is a predisposition to the good; there is no conceivable ground for us, therefore, from which moral evil could have first come to us" (6:43). In other

³⁰ I would therefore also oppose DiCenso's (2012) treatment of *Religion* insofar as he, too, tends to read the metaphysically sticky elements of the text in an "as if" spirit.

words, given that we are not originally evil but rather possess a pre-disposition to the good (which Kant explains in 6:26–28), it is not clear how to account for our free choice of evil. This concern is reflected in the quotation from Descartes given as an epigram: if it is up to us to determine freely our fundamental practical nature, why choose evil rather than the perfection of goodness? The problem here is actually worse, for if we *were* originally evil—evil by *nature* and not by choice—then our ontological guilt would not be imputable to us, and we would be back where we started, namely, with the problem of how to make sense of the imputability of guilt.

Earlier in the text, however, Kant gives voice to a different though related worry, which I think is already present in the quotation from *Rel.* 6:43. This worry will stand at the center of my interpretation of Schelling and Schopenhauer in chapters 2 and 3. The problem is this. As we saw, the adoption of the fundamental maxim—“the first subjective ground of the adoption of the maxims” (*Rel.* 6:25)—in which our original sin consists cannot just be a brute fact about our nature, for then all moral responsibility would dissolve into nothing. But how can we account for the choice of this fundamental maxim? It can be *either* arbitrary, and hence irrational, in which case, it seems, the entire edifice of moral responsibility will turn out to be groundless, which would undermine both moral responsibility and guilt (both ontological and factual), *or* the free choice of the maxim can be grounded in a deeper, even more fundamental maxim. But, Kant says,

there cannot be any further cognition of the subjective ground or the cause of this adoption (although we cannot avoid asking about it), for otherwise we would have to adduce still another maxim into which the disposition would have to be incorporated, and this maxim must in turn have its ground. (*Rel.* 6:25)

In other words, though we are inevitably drawn to ask this further question about the deeper ground for our adoption of our

fundamental maxim, we cannot cognize this ground. Why, exactly? Here Kant's answer seems to be that were it possible for us to cognize the ground, the question about *its* ground would arise, and so on and so forth, which would quickly confront us with an infinite regress. "Hence," Kant continues,

since we cannot derive this disposition, or rather its highest ground, from a first act of the power of choice in time, we call it a characteristic of the power of choice that pertains to it by nature (even though the disposition is in fact grounded in freedom) [*so nennen wir sie eine Beschaffenheit der Willkühr, die ihr (ob sie gleich in der That in der Freiheit gegründet ist) von Natur zukömmt*]. (Rel. 6:25)

It is hard to see what Kant is trying to claim here, but here is one possible interpretation: even though it must be the case that the disposition *is* grounded in freedom (otherwise, moral responsibility and, with it, guilt would evaporate), since we cannot really make sense of this act of grounding, given the problem of infinite regress, we are, so to speak, forced to consider—to "call"—the power of choice evil *by nature*. Kant sees that for moral guilt and moral responsibility in general to be saved, the agent must be responsible *all the way down*, for once we introduce an element that is supposed to account for the agent's immoral deeds and yet is one for which the agent *cannot* be deemed responsible, then this would get the agent off the hook, and the whole moral construction of guilt and blame would collapse. But since the idea of *all the way down* responsibility raises the specter of infinite regress, we find ourselves attributing an evil *nature* to the power of choice. The knot cannot be tied, and the cord of responsibility threatens to fray at its end, and this requires introducing talk of the choice's evil "nature" as a kind of stopgap and, thus, an introduction of determinism.

But this attribution of a natural and evil characteristic to the power of choice, it seems, is not expressive of Kant's considered

view. It seems, rather, that Kant here just wishes to explain why we tend to *talk* of man's corrupt "nature." This talk, however, Kant seems to think, is just a *faute de mieux* of sorts that is used to cover over our ignorance, our inability, that is, to square the circle, to block the regress *and* secure imputability.

But then the problem of infinite regress is left unaddressed: to block the regress, more is needed than merely to "*call* it a characteristic of the power of choice" and so on (*Rel.* 6:25; emphasis added). It thus seems that at the end of the day, *Kant cannot make sense of factual moral guilt* since he cannot make sense of moral responsibility, insofar as the latter is ultimately grounded in our free adoption of our ontological guilt.³¹ It seems that to block the regress, we must positively *assert* that this characteristic indeed belongs to the power of choice by nature. We must assert, differently put, that our original sin is both imputable to us (we are freely responsible for it) *and* attributable to our nature (it characterizes our power of choice, thus determinism).

But can this view be maintained? Can freedom and nature be united in this way, or is this idea offensive to reason no less than that of infinite regress, and as such does not provide us with an acceptable solution to the problem of the imputability of our ontological guilt and thus of guilt *überhaupt*? As we will see in the next chapters, Schelling and Schopenhauer³² address the Kantian problem and see that we have to think of ourselves as already determined in our nature in some morally significant way (to block the infinite regress) and yet as *freely* choosing our moral nature (to secure imputability). Thus, to avoid the regress, we would have to think of our being/nature as constituted by our free power of choice. Differently put, I want to suggest, for these thinkers, we

³¹ Saunders (2019) argues—convincingly, in my view—that Kant cannot make sense of our moral and legal ascriptions of *degrees* of responsibility. But here I argue that Kant cannot make sense of moral responsibility *simpliciter*.

³² Though, as I explain toward the end of the chapter on Schopenhauer, Schelling much more clearly adheres to the solution I am here adumbrating.

would have to posit the human being as *causa sui*—as a cause of himself or herself—where the human being freely chooses his or her own moral being, and this on the basis of this moral being or nature. Indeed, once the problem of moral guilt is pushed far enough, as we can see with Kant, it seems that the notion of *causa sui* is the only possible solution to the problem of accounting for the possibility of moral guilt.³³

Though Kant does not clearly endorse this idea, interestingly enough, before *Religion*, he can be seen to be making claims that take him close to the idea of the self as a cause of itself. He thus claims in the second *Critique* (I quote at some length):

the very same subject, being on the other side conscious of himself as a thing in itself, also views his existence *insofar as it does not stand under conditions of time* and himself as determinable only through laws that he gives himself by reason; and in this existence of his nothing is, for him, antecedent to the determination of this will, but every action—and in general every determination of his existence changing conformably with inner sense, even the whole sequence of his existence as a sensible being—is to be regarded in the consciousness of his intelligible existence as nothing but the consequence and never as the determining ground of his causality as a *noumenon*. So considered, a rational being can now rightly say of every unlawful action he performed that he could have omitted it even though as appearance it is sufficiently determined in the past and, so far, is inevitably necessary; for this action, with all the past which determines it, belongs to a single phenomenon of his character, *which he gives to himself* [*den er sich selbst verschafft*] and in accordance with which he imputes [*zurechnet*] to himself, as a cause independent of all sensibility, the causality of those appearances. (5:98; final emphasis added)

³³ Again, compatibilistic alternatives were deemed philosophically untenable by Kant.

Kant thus seems to say here that even though from the empirical perspective all of one's actions are determined, from the intelligible perspective they are all to be seen as a consequence of one's intelligible character, which the agent gives to himself and as a consequence becomes the locus of imputability for his actions, not least in his own eyes. This sounds very much like the idea of being a cause of oneself, being a *causa sui*.

Of course, a lot hangs here on what Kant means by an agent's intelligible character—a vexed issue, to say the least.³⁴ One way to think about one's intelligible character, suggested by Allison, is to understand it as spontaneity—the necessary presupposition for thinking about oneself and others as rational agents, where this involves the ability, given in apperception, to incorporate incentives (take them as reasons for action) and be conscious of this very activity of “taking as,” thus relating it to oneself (Allison 1990, 40–41). It appears, however, and this is what I assumed in my earlier discussion, that Kant has something more substantial in mind with the idea of intelligible character. As he writes in the first *Critique*: “Every efficient cause must have a character, that is, a law of its causality without which it would not be a cause” (A539/B567). It thus seems that at least part of what the intelligible character involves is not merely the self-conscious ability to incorporate incentives into one's maxim but precisely the maxim or the practical “law” *itself*, with reference to which incentives are incorporated and on the basis of which the agent acts.

If this is correct, then in the earlier quotation from the second *Critique*, Kant seems to be saying that an agent gives himself his own character, where the latter is to be understood as the “evil and unchangeable principles freely adopted” (5:100). While this is indeed a form of self-constitution, it does not amount to the

³⁴ As an indication of this, see Hermann Pistorius's admission, where he claims that he finds the empirical-intelligible character distinction to be “the most obscure and incomprehensible in the entire critical philosophy” (quoted in Allison 1990, 29).

full-blown assertion of the agent being a *causa sui*, since it does not make the claim that the agent gives himself his very own being or nature but merely his or her fundamental principles of action.³⁵

V. Conclusion

In this chapter, I argued that factual guilt for Kant is based in the final analysis on our ontological guilt, a guilt the imputability of which implies a free, intelligible deed. Thus, in Kant's philosophy, when pursued far enough, guilt for our immoral actions forces us to seek our guilt at a deeper level; it is only because we freely adopted our fundamental maxim that we can be guilty for our specific immoral deeds. It turns out on close inspection, however, that for this ontological guilt to be itself imputed to the subject, we must conceive of the subject as in some sense a cause of himself or herself. Kant, though he makes very similar-sounding claims, does not really embrace this idea, but, as we shall see, it is explicitly embraced in the thought of Schelling and Schopenhauer.

One final and significant issue should be raised. As I have reconstructed his argument, Kant derives our intelligible freedom from the imputability of our original sin, a freedom which can then be also used to anchor our factual guilt. A fatal problem—which I have so far put to the side—is that Kant does not provide good grounds for thinking that all of us are indeed ontologically guilty, that the human species is evil as such and “cannot be judged otherwise” (*Rel.* 6:32). To base this judgment, Kant appeals to anthropological anecdotes about the murderous and cruel behavior of certain “exotic” peoples (*Rel.* 6:33). But first, empirical data, reliable

³⁵ With his talk of giving oneself one's own character, Kant could not have in mind his concept of autonomy (unless one has a more copious concept of autonomy), for the latter involves giving to oneself laws of pure practical reason—the categorical demands of morality—while here Kant is also concerned with one's vicious constitution, which as such could not possibly be derived from pure practical reason.

as it might be (and it is probably not), cannot possibly ground the *universal* indictment of the human species Kant wishes to uphold. Second, even if empirical evidence demonstrated the absolute generality of such types of behavior, it would not give reason to consider this as evil in the moral sense, that is, as something for which we are responsible (cf. Allison 1990, 154).

One might wonder, then, whether it is possible to drop Kant's talk of species-wide, imputable radical evil and pursue the line of thought introduced in the second *Critique*, where, as we saw, *empirical guilt* is claimed to be justified or "founded" because it is "the consequence of the evil and unchangeable principles freely adopted" (5:100). That is, why not, in a manner analogous to the "fact of reason" argument (5:31; cf. A547–548/B575–576),³⁶ start with the assumption that in our experience of empirical guilt we are conscious of the feeling's validity (unless we were acting under duress, etc.) and then argue transcendently that a free, intelligible deed is that by virtue of which the feeling is justified? As we shall see, this is the spirit of the approach pursued by Schelling and Schopenhauer in their thinking about guilt.

³⁶ See Kleingeld 2010 for a reading of the "fact of reason."

2

Schelling

Evil, Freedom, and Guilt

Introduction

In this chapter, I examine Friedrich Wilhelm von Schelling's thought on the phenomenon of guilt. The main point I wish to establish in the following is that on the one hand, Schelling can be seen to follow in Kant's footsteps in that he, too, sees guilt as ultimately grounded in a free, intelligible act. On the other hand, Schelling goes beyond Kant in that he holds that for guilt (empirical or ontological) to be justified, it is necessary that the human being constitute or choose his or her own being in a timeless and free deed. In other words, Schelling endorses a version of the transcendental argument for freedom where guilt implies that the human being is a free *causa sui*. Schelling's conception of the human being as *causa sui* enables him to address the problem of infinite regress that, as we saw, beleaguers Kant's view.

While Schelling presents us with a rich and comprehensive philosophy, it is naturally not my aim here to explore all the details of his evolving thought but to merely extract what I think is essential to the problems at hand. As a result of this methodological limitation, I focus on a single work from his extensive output, namely, his 1809 *Philosophical Investigations into the Essence of Human Freedom*. This narrowing down of my discussion is justified, I believe, precisely because it is in this work—rather than his works on the philosophy of nature, art, or history—that his belonging to

the Kantian, metaphysical tradition with respect to guilt is clearly apparent.

I first present the essential contours of his *Investigations*, focusing on the elements most relevant to my problem, and then, in the second part, I turn to examine Schelling's thinking about the free act of self-creation that lies at the basis of personal responsibility and guilt.

I. Schelling's Idealistic System in Outline

In his *Philosophical Investigations into the Essence of Human Freedom*, Schelling, as part of an attempt to address the theological problem of evil, engages, among others, with Kant's views on evil and proposes a solution to the question of its origin.¹ For Schelling, evil or radical evil is not simply a matter of reversing one's own subjective maxims and putting one's own self-interest above the moral law. This way of looking at evil, according to Schelling, amounts to a failure to take evil sufficiently seriously. Rather, evil for Schelling arises from the depth of being itself, so that the possibility for evil—albeit not its actuality—lies in God himself. And it is precisely this way of looking at evil—as arising out of the heart of being as such—that enables Schelling potentially to succeed where Kant failed: whereas for Kant, as we saw, the origin of evil remains “inexplicable,” Schelling attempts to illuminate it on the basis of fundamental metaphysical considerations.

The problem of evil for Schelling stands at the heart of a number of related issues he seeks to think through and illuminate in his treatise. Specifically, the central issue is how to make sense of *pantheism* in a manner that makes room for *freedom* and *evil* and

¹ I was initially made aware of the connection between Kant's and Schelling's view on the question of evil and guilt by Žižek 2009, 185–190. I will take issue with some of the points he makes.

yet *justifies God* in the face of this evil, thus addressing the traditional problem of *theodicy*. The philosophical challenge here is immense, not only because of the fundamental nature of each of the highlighted notions but also because there is deep tension between these ideas when thought together, what makes it difficult to integrate them systematically. At the nucleus of the tension stands the idea of pantheism, the being of all in God.² First, if all is in God, then, it seems—especially if we think about pantheism in a Spinozistic fashion—that all events are interconnected in a systematic fashion, where there is no contingency or arbitrariness and where what Schelling also calls “fatalism” (e.g., *Phil. Inv.* 11) rules. But if that is the case, then how to think the reality of freedom? Indeed, in Baruch Spinoza’s philosophy, thorough determinism obtains, and free will is an illusion. Pantheism also proves challenging for the idea of evil and the possibility of vindicating God in the face of it. This is a familiar problem: if all is in God, and God is absolutely Good, then how can the reality of evil be explained in such a way that does not render God in some sense responsible for it, which would contradict, if not his absolute benevolence, then his omnipotence or omniscience?

Two possible ways to escape this predicament come to mind. One way is to argue that because of the fatalism that characterizes pantheism, freedom does *not* exist, and given that it requires freedom, evil is also nonexistent. The other option would be to reject necessity, which Schelling also calls reason (e.g., *Phil. Inv.* 11), and uphold freedom. But Schelling asserts that neither option is open to us:

To pull oneself out of the conflict by renouncing reason seems closer to flight than to victory. With the same justification,

² Schelling distinguishes a number of different ways to understand this term (*Philosophical Investigations into the Essence of Human Freedom* [hereafter *Phil. Inv.*], 12–15) but seems to settle on the one I formulate in the body of the text with the caveat that it be sufficiently “brought to life and torn from abstraction” (*Phil. Inv.* 20).

another could turn his back on freedom in order to throw himself into the arms of reason and necessity without there being cause for triumph on either the one or the other side. (*Phil. Inv.* 11)

And why exactly is renouncing either not really triumph but defeat? Schelling's view is that such renunciation would be tantamount to an escape from real philosophy: "connection of the concept of freedom with the whole of a worldview will likely always remain the object of a necessary task without whose resolution the concept of freedom would teeter while philosophy would be fully without value" (*Phil. Inv.* 10). Since this lies beyond the scope of my interests here, I will not pursue the question of what reason Schelling possesses for adhering in the first place to pantheism (= the existence of all in God = necessity, fatalism, determinism) but would like to raise it with respect to freedom: namely, why does Schelling insist that freedom is real at all? Schelling can be seen to address this question right at the beginning of his essay, when he claims that

Philosophical investigations into the essence of human freedom can in part address the correct concept of freedom in so far as the fact of freedom, no matter how immediately the feeling of which is imprinted in every individual, lies in no way so fully on the surface that, in order merely to express it in words, more than an uncommon clarity and depth of mind would not be required; in part, they can deal with the connection of this concept with the whole of a scientific worldview. (*Phil. Inv.* 9; translation slightly amended)

Two points are worth emphasizing here. First, Schelling here lays at the basis of his discussion the reality of freedom: freedom is a fact (*Thatsache*), one that is given to us immediately (*unmittelbar*) as a *feeling* (*Gefühl*). And though the feeling itself by itself is not enough to explicate the nature of freedom or its place within pantheism—this requires the development of a "whole scientific

worldview”—the fact of freedom as given in immediate feeling is sufficient as a starting point for philosophizing. Indeed, the reality of freedom is absolutely fundamental to Schelling’s idealism as such and could be without great exaggeration identified with it: idealism for Schelling is the “thought of making freedom the one and all of philosophy” (*Phil. Inv.* 22).³ Accordingly, since freedom is a fact, room must be found for it in nature itself: “everything real . . . has activity, life and freedom as its ground” (22). This is Schelling’s “higher realism,” where Kant’s distinction between spirit/freedom and nature is transcended insofar as nature itself becomes infused with freedom and freedom is “spread . . . throughout the whole universe” (22). What feeling of freedom, however, does Schelling have in mind here? He does not say, but, as I will argue, the feeling of freedom Schelling mentions here is the feeling of the free act of one’s own self-creation, a freedom that comes to the fore and impresses itself most forcefully in the experience of *guilt*.

The second point to note about this opening paragraph is Schelling’s proclamation that his investigations into human freedom would show how to think freedom in a way that does not shrink back in the face of the tensions we have noted but would explain how freedom can find its place within the systematic whole, within the pantheistic Absolute. As a matter of fact, only a couple of pages later, Schelling even makes the strong claim that “many are brought to this viewpoint [of pantheism] through the most lively feeling of freedom” (11). In other words—and this is something that Schelling will try to show—not only is pantheism not in contradiction with freedom, but it actually follows from it, as can be appreciated once one’s concepts of freedom and pantheism are sufficiently developed and understood. For Schelling, then, the feeling of freedom is the guiding thread for the erection of the system. This is important because it shows that the entire metaphysics Schelling

³ See also the centrality of freedom in the inaugural “Oldest System Program of German Idealism,” co-composed by Hegel, Schelling, and Friedrich Hölderlin.

offers is motivated by the need to make room for freedom and thus for moral responsibility. What does Schelling mean by freedom, however?

In his close reading of Schelling's *Philosophical Investigations*, Heidegger usefully distinguishes between several concepts of freedom so as to highlight Schelling's contribution to the thinking of freedom:

1. Freedom as capability of self-beginning.
2. Freedom as not being bound to anything, freedom *from* (negative freedom).
3. Freedom as binding oneself to, *libertas determinationis*, freedom *for* (positive freedom).
4. Freedom as control over the senses (inappropriate freedom).
5. Freedom as self-determination in terms of one's own essential law (appropriate freedom), formal concept of freedom. *This includes all the previous determinations.* (Heidegger 1985, 88)

In Heidegger's view, the last concept, that which retains and contains all the first four conceptions, is that of Kantian freedom. What is *Schelling's* concept of freedom, then, and how does it go beyond the Kantian one? Schelling claims: "the real and vital concept is that freedom is the capacity for good and evil," and he adds that "This is the point of most profound difficulty in the entire doctrine of freedom" (*Phil. Inv.* 23). As we shall see in more detail, this concept goes beyond the Kantian one in that while the latter focuses on a subject's self-determination in the free adoption of a fundamental maxim, Schelling's has to do with one's *free choice of one's very being*—with one's being good and evil. Another way to think about this difference is that while for Kant (moral) freedom involves the overcoming of nature, in Schelling freedom involves the free choice of one's own nature: freedom and nature are to be thought together.⁴

⁴ As Heidegger puts it, in Kant, nature "does not become constitutive for an independent ground of the whole existence of man. But where nature is thus understood, not

Schelling's concept of freedom thus involves the idea of *causa sui* in an unabashed form, as we shall see in more detail.

A quick note on Schelling's locution "good *and* evil" (as opposed to "good *or* evil"): this phrasing, as Heidegger explains, is not accidental. Schelling's idea here is that choice of oneself as evil involves at the same time "an essential relation of evil to good" (Heidegger 1985, 156), and vice versa. In other words, for Schelling, evil can only *be* what it is in relation to the good, specifically, in its strife against the good, and, conversely, the good can only *be* what it is in its struggle against evil.

A foundational building block that Schelling places at the beginning of his discussion is the idea of the *will* as primal being (*Ursein*). He clarifies that to the will "alone all predicates of Being apply: groundlessness, eternality, independence from time, self-affirmation" (*Phil. Inv.* 21). The idea that primal being is will implies that being at its very core is not a static substance but a living, dynamic striving or longing, a movement of *becoming*. Moreover, this dynamic, evolving longing, which Schelling also calls life, by virtue of its groundlessness, is at its basis free, and as free—and here Schelling is operating with Kantian ideas—is independent of time and is in this sense eternal. Finally, the will is characterized by self-affirmation—the will, in its striving, wills itself, seeks to actualize itself.

Being as will, however, is not one or unitary but envelops a distinction. Schelling thus claims: "The natural philosophy of our time has first advanced in science the distinction between being in so far as it exists and being in so far as it is merely the ground of existence" (*Phil. Inv.* 27). Being, in other words, is existence as well as ground—the latter also referred to occasionally as "nature"—where the ground is the ground *of* existence and where the ground *is* a ground only in relation to the existence which it grounds. In other

as what is merely to be overcome, but as what is constitutive, it joins a higher unity with freedom" (Heidegger 1985, 84).

words, the relation between the two is to be understood dialectically, where one is what it is only in relation to the other.

But—we may ask—if ground becomes ground only once it grounds existence, then what is there before ground? Toward the end of his treatise, Schelling therefore claims that “there must be a being *before* all ground and before all that exists, thus generally before any duality—how can we call it anything other than the original ground or the *non-ground* [*Ungrund*]?” And he continues: “Since it precedes all opposites, these cannot be distinguishable in it nor can they be present in any way. Therefore, it cannot be described as the identity of opposites; it can only be described as the absolute *indifference* of both” (*Phil. Inv.* 68). It is only once the non-ground “divides itself” (70) into ground and existence that there emerges the duality that makes up the life, the movement, of the whole.⁵

How to think of God in relation to this understanding of being as ground and existence? According to pantheism, all is in God. Does this therefore mean that God is both ground and existence? Schelling answers yes and no. The ground grounds God’s existence and is, moreover, indeed part of God, is *in* God—because there is nothing outside of God—and yet is *not* God. It is that in God which is not God. As Schelling claims:

Since nothing is prior to, or outside of, God, he must have the ground of his existence in himself. . . . This ground of his existence, which God has in himself, is not God considered absolutely, that is, in so far as he exists; for it is only the ground of his existence. It [the ground] is *nature*—in God, a being indeed inseparable, yet still distinct from him. (*Phil. Inv.* 27)

⁵ Schelling, however, does not explain how this division, this initial splitting of the non-ground into ground and existence, occurs. What is the “spark” that starts the whole movement going, that stirs the non-ground out of its indifference?

God thus has the ground of his existence in himself, and yet this ground is not God but distinct from God. It is this important idea that enables Schelling to address one aspect of the problem of evil, namely, the (seeming) evil of God: if all is in God (pantheism) and if evil—perpetrated by individual human beings—is real, then God, it would follow, is at least partly evil (contains evil), which contradicts his infinite goodness. The well-known Augustinian solution is to posit the free will of man: since human beings are *free*, they freely choose to do evil and thus bear this responsibility themselves, in a way that absolves God. But then the question becomes how to conceptualize this freedom: how could human beings be free and have a “root independent of God” (Heidegger 1985, 103) if they are in God (since all is in God)? In response, Schelling claims that for finite beings such as the human being,

to be divided from God, they must become in a ground different from God. Since, however, nothing indeed can be outside of God, this contradiction can only be resolved by things having their ground in that which in God himself is not *He Himself*, that is, in that which is the ground of his existence. (*Phil. Inv.* 28)

In this way, the independence of created beings from God is secured without abandoning the pantheistic axiom that all is in God. In this way, the reality of evil and God’s infinite goodness are explained as compatible. It is this way of understanding pantheism—as all being in God without everything being identical to God—that opens for Schelling a way out of this traditional problem.⁶

⁶ I am not interested here in providing a thorough assessment of Schelling’s solution to the problem of evil. Let me just say that the solution sketched here on Schelling’s behalf amounts to a significant concession: God is not omnipotent, for, on the one hand, he needs the ground—the possibility of evil—in order to exist, and, on the other hand, he seems to lack the power to prevent evil from actualizing itself and becoming real. We will return for more on evil.

The emergence of God from out of the ground which is in him as well as the coming to be of created beings, including the human being, must now be explained. According to Schelling, the coming to be of God can be seen as a becoming in which God is “begotten in God himself” (*Phil. Inv.* 30)—God is begotten in that which is in him but is not him, that is, the ground. God thus delivers himself from out of that which is in himself but is not identical to himself. This self-realization of God is expressive of the will in God, of the longing in God, to become himself from out of the ground which is in him: “it is the yearning the eternal One feels to give birth to itself” (28). There is in God a longing to reveal himself to himself from out of that which is in himself. And it is in the process of revealing himself that God also creates the world. The coming to be of the world is the becoming self of God, the revealing of God to himself. God reveals himself to himself by creating his other, by creating that which is not-him.⁷

The yearning of God—the expression of will in the divine—is, however, only one aspect of the being which is will, since the ground itself also shares in this willing and yearning. Thus, Schelling claims that “The yearning is not the One [God] itself but is after all co-eternal with it” (*Phil. Inv.* 28). This yearning of the ground participates in the process of giving birth to God (28), but, as ground, and in contrast to God, it lacks all understanding, even though it desires the emergence of the understanding of God. The ground is *dark*—it lacks all consciousness and understanding and thus all reason and *light*—but it nevertheless intimates the light of God; it is “not a conscious but a divining [or intimating] will” (29). It strives to come to the light by participating in the process of God’s becoming out of it itself. But this means that in some sense, the ground already includes God within itself, *in*

⁷ See Marx 1984 for further detail regarding the emergence of God and his relation to evil; and Žižek 1996 for the view that the ground from which evil arises is the refuse God leaves behind, so to speak, when he enters into being, which in my view is not quite right, since God also emerges *out* of this ground.

potentia: "Because . . . this being (of primordial nature) is nothing else than the eternal ground for the existence of God, it must contain within itself, although locked up, the essence of God as a resplendent glimpse of life in the darkness of the depths" (30). On the other hand, and at the same time, the ground is also "aroused by the understanding" (30): the God that emerges from the ground also stirs the ground to affirm its willing itself into God. Nevertheless—in a further twist of this dialectical relation between ground and existence, nature and God—the more the ground gives rise to God from out of itself, the more it strives to remain ground, that is, "to close itself up in itself so that a ground may always remain" (31).

As claimed earlier, in the process of God's striving to become, the world of beings is created. This occurs, according to Schelling, in the introduction into the ground of order, reason, and unity—which Schelling calls "the word" of God (32)—which creates the entire kingdom of nature, both inanimate and animate, in its lawfulness and in its inner (the internal purposefulness and ordering of the parts that make up the individual) as well as its outer (species and genera) organization. This is accomplished by a process Schelling calls "division." He thus claims that "The first effect of the understanding in nature is the division of forces, since only thus can the understanding unfold the unity that is unconsciously but necessarily immanent in nature as in a seed" (30). In the ground lies anarchy, darkness, but also potential for ordering, for organization. This process of "division," of ordering, thus brings about a synthesis, a ligament or bond, between the two principles of ground and existence, also referred to by Schelling as the twin principles of darkness and light, anarchy and order, non-understanding and understanding, creaturely self-will and universal will. Still, no matter how thoroughgoing the process of division is, no matter how deep into the "center" of the ground it goes and how high to the peaks of light it ascends, there will always remain the ground, that which resists division, which Schelling calls the "incomprehensible base of reality in things . . . the indivisible remainder" (29). Thus, "After

the eternal act of self-revelation, everything in the world is, as we see it now, rule, order and form; but anarchy still lies in the ground, as if it could break through once again" (29). So now we can see more clearly the contrast between, on the one hand, the will of God, which is "to universalize everything, to raise everything up toward unity with the light or keep it there," and, on the other hand, "the will of the ground . . . to particularize everything or to make it creaturely" (47).

The movement of creation, as it evolves, involves an ever deeper and ever more complex division of forces which at once both lifts the ground into a higher and higher light and burrows into the darkest recesses of the ground. The more it continues in this manner and the more complex the creature created, the more its being is constituted by the brightest of lights as well as by the deepest depths of the ground. The movement here can be compared to the movement of a sort of twisting corkscrew (compare the illustration in Heidegger 1985, 136), where the deeper one goes, the higher one gets; the deeper one descends into the dark ground, the more light emerges and the higher the ground is raised into the understanding, into the universal. In the case of *man*, the twist of the "corkscrew" reaches the deepest and the highest, which is why in man, there is a union of the darkest ground of all and the brightest light. As Schelling puts it:

This raising of the deepest *centrum* into light occurs in none of the creatures visible to us other than man. In man there is the whole power of the dark principle and at the same time the whole strength of the light. In him there is the deepest abyss and the loftiest sky or both *centra*. (*Phil. Inv.* 32)

In man, the deepest darkness but also the brightest light are united. Moreover, for Schelling, and precisely for this reason, man is the crown of creation: in man, "the word is fully proclaimed which in

all other things is held back and incomplete" (32). And with the stepping of man onto the stage of creation, we turn to consider the questions of evil and guilt.

II. Freedom, Evil, and Guilt

As we saw, Schelling's metaphysics posits the twin principles of ground and existence, darkness and light, self-will and universal will, and explains the coming to be of the world of created beings as the bonding or tying together of these two principles. It is with this emergence of creaturely beings that God reveals himself. Moreover, in God, the coming to be of which is the process of the bonding of the principles, there is an essential and eternal unity between ground and existence, a unity Schelling calls the "eternal spirit" (*Phil. Inv.* 32). Now, in a striking dialectical move, Schelling makes the claim that for God to *be*, that is, reveal himself as he is, he needs an Other in which the two principles of darkness and light are *not* indissolubly united; specifically, to exist, God needs man. Here is the relevant passage:

[S]pirit, that is, *God* as existing *actu*, reveals itself in the proclaimed word. In so far as the soul is now the living identity of both principles, it is spirit; and spirit is in God. Were now the identity of both principles in the spirit of man exactly as indissoluble as in God, then there would be no distinction, that is, God as spirit would not be revealed. The same unity that is inseparable in God must therefore be severable in man—and this is the possibility of good and evil. (32–33)

In other words, not only does God create man and thus make man possible, but the converse is also the case: man makes *God* possible. Thus, not only is there no contradiction between his existence and the possibility of (good and) evil, but the two are co-conditioned.

To further clarify what is at issue here, we need to take a closer look at what Schelling understands by “spirit.”

First, spirit not only involves the unity (indissoluble in God, severable in man) of the most central ground with the highest light, but it also accomplishes the raising “from the creaturely into what is above the creaturely” (*Phil. Inv.* 33). The human being, as spiritual, though a created being, does not belong to what is creaturely, that is, to nature: the spiritual is “above and outside of all nature” (33). Second, spirit “is will that beholds itself” (33): spirit involves self-consciousness and with that a transcendence of itself. In Hegelian terms, the in-itself of the union of principles becomes in man a for-itself. Third, precisely in this self-consciousness and transcendence of nature lies “complete freedom”: spirit is “free from both principles” (33) and can thus rise above both ground and light. More concretely and more important for our purposes, the spirit in man entails the ability to undo the bond of ground and existence where the former is subordinated to the latter and *reverse* their relation in such a way that ground, or creaturely self-will, usurps the principle of light or universal will in the human being and subjects it to its rule—and this is precisely what evil is. As Schelling puts it:

[T]hat precisely this elevation of self-will is evil is clarified by the following. The will that steps out from its being beyond nature, in order as general will to make itself at once particular and creaturely, strives to reverse the [proper] relation of the principles, to elevate the ground. (34)

Evil thus arises when a reversal of the proper relation between the two basic principles is effected: instead of subordinating the particular creaturely will to the universal will, man separates the principles and reattaches them in a reversed fashion, thus making “the spiritual within himself into a means” (54) of his dark self-will. Note, however, that this reversal, since it presupposes the highest

light in creation, involves not a blind willing and striving: the ground elevates itself above the principle of light *by means of spirit*; darkness becomes self-aware and posits itself above the light. In other words, the reversal is guided by the ground's spiritual self-consciousness; the act of reversal is done in full awareness—the evil person, in a way, *knows what he or she is doing* and is thus morally culpable. This reversal of ground and existence, of self-will and universal will, involves man's fall from

truth into lies, from light into darkness, in order to become a self-creating ground and, with the power of the *centrum* which he has within himself, to rule over all things. . . . From this arises the hunger of selfishness which, to the degree that it renounces the whole and unity, becomes ever more desolate, poorer, but precisely for that reason greedier, hungrier, and more venomous. In evil there is the self-consuming and always annihilating contradiction that it strives to become creaturely just by annihilating the bond of creaturely existence, and, out of overweening pride to be all things. (55)

The evil person, as spiritual, thus uses his powers of understanding (the brightest light) to impose his individuality, his self-will or selfishness, on all things so that they serve his individuality. As Schelling claims, this form of being involves a "contradiction" where unity and reason, instead of maintaining and further expanding themselves, are used to undo themselves—undo the bond—for the sake of self-will.

We saw that spirit for Schelling means freedom and that man's freedom, as a spiritual being, is the freedom for good and evil—a freedom that consists in the ability to separate and reverse the basic principles. To deepen our grasp of Schelling's notion of freedom and to see how guilt enters his view, further elements have to be introduced. The first thing to point out is that Schelling adheres both to the view that an indeterminist conception of freedom is

untenable and to the truth of empirical determinism. Thus, with respect to the former, he claims that

the common concept of freedom, according to which freedom is posited as wholly undetermined capacity to will one or the other of two contradictory opposites, without determining reasons but simply because it is willed . . . when applied to individual actions, it leads to the greatest inconsistencies. To be able to decide A or –A without any compelling reasons would be, to tell the truth, only a prerogative to act entirely irrationally. (48)

Schelling's rejection of indeterminist freedom—a close cousin of leeway freedom (McKenna and Pereboom 2016), the ability to act otherwise—as irrational is especially important in relation to guilt: an entirely irrational action performed by an entirely irrational agent who just wills the action without any rhyme or reason whatsoever is arguably not an action for which there is an agent at all and thus no culpable agent who should feel guilty or be held guilty by others for this action.

On the other hand, Schelling, in a Kantian fashion, accepts the idea that determinism rules the realm of phenomena and claims that

Determinism (or, according to Kant, predeterminism)⁸ counters this system of the equilibrium of free will and, indeed, with complete justification, since it claims the empirical necessity of all actions because each is determined by representations or other causes that lie in the past and that no longer remain within our power during the action itself. (49)

⁸ As the translators of Schelling's *Philosophical Investigations* note in their footnote, Schelling is here referring to Kant's footnote in *Religion* (6:50).

Though Schelling accepts determinism at the level of the phenomena, he does not regard determinism as the final word with respect to the question of freedom. He therefore formulates an idea of freedom that proceeds from a “higher” point of view, a point of view that upholds “a higher necessity . . . which is equidistant from contingency and from compulsion or external determination, which is rather an inner necessity springing from the essence of the acting individual itself” (49). In other words, Schelling posits a freedom that is *one* with nature or determinism,⁹ thus potentially blocking the Kantian infinite regress—a determinism, however, that is not located at the empirical level, which would deprive the human being of freedom given the “empirical necessity of all actions” (49), but at the *intelligible* level. Here Schelling makes use of an idea that can be found in Kant but which he makes his own: the idea that individual human beings have an *intelligible* being—in Kant, we have the notion of intelligible *character*—which lies outside of space and time, the realm of phenomena where strict, empirical determinism reigns.

Let’s first approach Schelling’s notion of intelligible being. Schelling writes that his idealism

first raised the doctrine of freedom to that very region where it is alone comprehensible. According to idealism, the intelligible being of every thing and especially of man is outside all causal connectedness as it is outside or above all time. Hence it can never be determined by any sort of prior thing since, rather, it itself precedes all else that is or becomes within it, not so much temporally as conceptually. . . . We are expressing namely the

⁹ According to Michelle Kosch, Schelling’s *Philosophical Investigations* forms a transitional work with respect to his conception of freedom, one in which he endorses both a form of compatibilism and a form of indeterminism—a tension which, in her view, accounts for some of the inconsistencies in the text (Kosch 2014, 145–146). On my reading of Schelling’s conception of freedom, however, indeterminism is not part of the story at all; one is determined in one’s intelligible being through and through—determined by oneself.

Kantian concept not exactly in his very words, but indeed in the way, as we believe, that it would have to be expressed to be comprehensible. (49)

How does Schelling's idea of an intelligible being differ from the Kantian notion? As we saw in chapter 1, the Kantian notion of the intelligible character involves not merely the self-conscious ability to incorporate incentives into one's maxim but precisely the maxim or the practical "law" *itself* with reference to which incentives are incorporated and on the basis of which the agent acts. Schelling's view, in contrast, is concerned not with maxims or laws that a person adopts but with a person's intelligible *being* itself, with the "essence of the acting individual itself" (49). And since *will* is primal *being* (21), and "many individual wills are included in a primal will" (23), the intelligible being, the essence, of an individual person is constituted by the relationship between the two aspects of the will corresponding to the two basic metaphysical principles: creaturely will (or self-will) and universal will. Now, given that freedom for good and evil consists in the ability to reorder the principles, this freedom consists in the ability *to constitute one's very being*, one's very intelligible self or essence, which lies outside of space and time. How, however, is this freedom at the same time a form of necessity, of determinism? Here is the key passage:

In order to be able to determine itself [the intelligible being] would already have to be determined in itself, admittedly not from the outside which contradicts its nature, also not from the inside through some sort of merely contingent or empirical necessity . . . but rather it would have to be determined by itself as its essence, that is, as its own nature. This is of course not an undetermined generality, but rather determines the intelligible being of this individual. . . . But what then is this inner necessity of the being itself? Here lies the point at which necessity and freedom must be unified if they are at all capable of unification. Were this

being a dead sort of being and a merely given one with respect to man, then, because all action resulting from it could do so only with necessity, responsibility and all freedom would be abolished. But precisely this inner necessity is itself freedom; the essence of man is fundamentally *his own act* [*That*]. . . . [I]t is a primal and fundamental willing, which makes itself into something and is the ground of all ways of being. (49–50)

Man chooses his own intelligible essence in a free act of self-creation that is not arbitrary but rather follows with necessity *from* this, his own intelligible essence. The human being is thus his own necessary cause or *causa sui*. Moreover, being an intelligible and necessary *causa sui* is in Schelling's lights not only *compatible* with freedom and responsibility, but, stronger still, it is "itself freedom" (49–50), as he puts it. Though man is an "undecided being" (51), "only man himself can decide" (51) and constitute himself according to his own essence.¹⁰ And it is precisely this free act or deed (*That*) of self-constitution whereby the person "contracts" his "radical evil" (53) and subverts the universal principle in favor of the dark ground of animality. Further, and in a similar fashion to Kant's choice of one's fundamental maxim, the deed "does not itself belong to time but rather to eternity" (51). As we saw, Kant had a hard time making room for both freedom and determination: the former is needed for moral imputability, while the latter is required to avoid sheer arbitrariness (and thus irrationality) as well as the infinite

¹⁰ With respect to this, Kosch writes that "the intelligible agent does not choose among different options, since that would require preexisting its own activity as self-constitution, and that is incoherent" (Kosch 2014, 150). I agree that Schelling's concept of self-creation does not make room for the idea of leeway freedom or alternative courses of action, but it seems the problem Kosch raises here is not with the idea of alternative possibilities as such but with the very idea of self-constitution which requires one to pre-exist one's act of self-constitution. This is, however, for better or worse, Schelling's view, and we will encounter it again in the discussion of Schopenhauer.

regress he catches a glimpse of in *Religion*. Schelling's solution is to combine the two in a conception of freedom that just *is* necessary (self-)determination.

We should highlight the transcendental argument that Schelling is putting forward at the end of the long passage quoted earlier (*Phil. Inv.* 49–50), for it is of a structure we are already familiar with from our discussion of Kant: in contrast to “dead” necessity which excludes freedom, an “inner necessity” which is “itself freedom” is presupposed, for otherwise, “responsibility and all freedom would be abolished” (50). In other words, for our moral responsibility for our actions—a necessary condition of our empirical guilt—to be attributable to us, we have to be considered as having actually exercised our freedom—performed a deed—at the intelligible level in the constitution of who we are. The justifiability of empirical guilt thus implies intelligible freedom.

Another point of importance is the following. The freedom Schelling posits at the basis of our self-constitution, the freedom that is compatible with necessity and is the only true freedom according to Schelling, is a peculiar kind of freedom since it does not involve leeway freedom, the possibility of having acted differently. Since one's self-creation is necessitated by who one is, it is impossible for one to have acted and created oneself differently. If one is *causa sui*, one could not really have been any different. Moral guilt and responsibility are secured, and room is made for an act of freedom, but the freedom at stake here is thus in no way the kind of leeway freedom often posited at the basis of guilt and responsibility.

Let us now look a little closer at how Schelling interprets the idea of original sin. For Kant, we saw, all human beings are guilty of original sin. Schelling claims that “all who are born are born with the dark principle of evil within” (*Phil. Inv.* 53), in such a way that although this original evil is “wholly independent of freedom in relation to contemporary empirical life, [it] is still in its origin [man's] own act and for that reason alone original sin” (53)—“every creature

falls due to its own guilt" (48).¹¹ Despite these words and Schelling's explicit invocation of the term "original sin," it is less than clear that he subscribes to the view that *all* human beings are guilty as such and are responsible for the same ontological guilt. Rather, it seems that he allows for the possibility that while some indeed freely determine themselves as evil, this is by no means the case for everyone, since for every person "the allowing-to-act-within-himself of [both] the good *and* evil principles" is "the result of an intelligible act whereby his being and life are determined" (54; emphasis added; see also 48). Thus, Dale Snow claims that it is the *temptation* to become evil that is "a defining characteristic of the human condition" (Snow 1996, 168), not the actual choice of evil. This temptation or arousal to evil is connected in Schelling's system to an idea we have encountered, namely, that the ground recedes into the darkness the more the light advances and spreads over the whole. As Schelling puts it specifically with respect to evil: "As a thunderstorm is caused in a mediated way by the sun but immediately by an opposing force of the earth, so is the spirit of evil . . . aroused by the approach of the good" (*Phil. Inv.* 46). Whether or not all are guilty of original sin, however, Schelling, like Kant, holds that this act of self-positing determines one's "being and life" (54): one's empirical actions are determined by one's original and timeless deed. One's original self-constitution thus *explains* one's deeds as well as one's empirical guilt.

A "free and eternal beginning" (51) thus lies at the basis of who we are, of our radical evil (in case we choose the dark principle over that of light), in a way that determines our empirical actions.

¹¹ The passage continues with the following words: "But just how in each individual the decision for good or evil might now proceed—this is still shrouded in complete darkness" (48). On the basis of this claim, Bernstein holds that "Like Kant, Schelling affirms that, in the final analysis, it is inscrutable why some beings choose to do evil and others choose to do good" (Bernstein 2002, 93). But Bernstein omits the remainder of the passage, where Schelling writes that this decision "seems to demand a specific investigation" (*Phil. Inv.* 48), which he then proceeds to provide on the basis of the idea of self-constitution. Now, whether, at the end of the day, Schelling's analysis is satisfactory is, of course, a different question.

Schelling is nevertheless aware of the radical nature of this idea of freedom:

As incomprehensible as this idea may appear to conventional ways of thinking, there is indeed in each man a *feeling* in accord with it as if he had been what he is already from all eternity and had by no means become so first in time. Hence, notwithstanding the undeniable necessity of all actions and, although each individual, if he is aware of himself, must admit that he is by no means arbitrarily or by accident good or evil, an evil individual, for example, surely appears to himself not in the least compelled . . . but rather performs his actions in accordance with and not against his will. That Judas became a betrayer of Christ, neither he nor any other creature could change, and nevertheless he betrayed Christ not under compulsion but willingly and with complete freedom. (51; emphasis added)

This crucial passage is packed with several suggestive claims. It seems that Schelling claims here that a glimpse of the original and eternal act of self-constitution is accessible to the agent herself in the form of *feeling*. Specifically, this intimation of intelligible freedom explains in Schelling's view the fact that we appear not compelled but free to ourselves in our empirical actions, even though we may know that, as empirical, they are in truth determined. And though Schelling admits that one cannot become fully conscious of this free and intelligible act of self-constitution, since "it precedes consciousness just as it precedes essence, indeed, first *produces* it" (51–52), a clearer consciousness of it is actual in the form of the scoundrel's admission when he claims pursuant to a commission of an immoral deed, "that's just the way I am," which Schelling interprets as amounting to the agent confessing that he is "*surely aware that he is like he is through his guilt*, as much as he is right that it was impossible for him to act otherwise" (52; emphasis added). In other words, I wish to suggest, it is precisely in

the feeling of empirical guilt that one most clearly recognizes one's intelligible deed of self-constitution and thus one's original or ontological guilt, according to Schelling.

And with the concept of feeling, we are back to the treatise's point of departure. As we saw, Schelling holds that the "fact of freedom" on which his entire system is based is given in feeling. It now becomes clear that this feeling is the feeling of our own eternal deed—a feeling that becomes especially acute when we judge ourselves to be guilty for an immoral deed. To put it differently, while we have seen that Schelling transcendently argues for intelligible freedom on the basis of "responsibility and all freedom" (50), here we can see him arguing for this freedom on the basis of the feeling of empirical guilt itself: to make sense of the freedom encountered in this feeling, we have to posit it at the intelligible level. Moreover, for Schelling, when we experience empirical guilt, the true object of this guilt is not the empirical misdeed itself but rather the eternal act through which we acquired our ontological guilt; that is what one is "aware" of, that is what one really feels guilty about. The phenomenal consciousness of one's own guilt discloses the metaphysical truth and attests, though dimly, to the free necessity at the depth of one's own being.

Slavoj Žižek interprets Schelling's claims about our vague awareness of our responsibility for who we are as providing us with "a very delicate psychological observation: sometimes we feel responsible without having effectively sinned; we feel guilty without accomplishing the act. This sentiment is, of course, the so-called sentiment of "irrational," unfounded guilt" (Žižek 2009, 189). But Schelling is actually trying here to give an interpretation of the phenomenology of quite ordinary empirical guilt. According to Schelling's analysis, when we feel guilty for an action, we find—and there is—nothing irrational about it. Rather, we experience our empirical guilt as arising out of who we are and yet somehow experience this nature of ours as of our own free making.¹² In

¹² In his reading, Žižek also emphasizes Schelling's view—anticipating psychoanalysis in some respects, according to Žižek—that our original and free choice of being in virtue

contrast to Žižek's reading, we are guilty and feel so precisely for "accomplishing the act"—the act of self-constitution. A similar insight governs our ascriptions of guilt to others: while we can foresee their behavior given the determinism that governs the empirical realm, we nevertheless are "convinced of [an] individual's guilt as [we] could only ever be if each particular action had stood within his power" (*Phil. Inv.* 52).

Before I conclude, a subtle but crucial difference between Kant's and Schelling's arguments must be stressed: while, as we saw, Kant argues for intelligible freedom on the basis of our ontological guilt, our original sin of which we are all guilty, Schelling can be seen to base his argument on *factical* guilt; it is our factical guilt, not our original sin, that implies our intelligible freedom. This is an important difference, because in basing his argument for freedom on factical guilt, Schelling avoids the flaw that threatens to ruin Kant's entire argument, namely, Kant's unproven assumption that the human species is indeed ontologically guilty. Instead of assuming that we are guilty of original sin, Schelling assumes that our empirical or factical guilt is genuine, which is a far more plausible assumption to make given that the feeling of guilt is immediately given in first-person consciousness as a fact. Critical pressure, however, as we shall see, can be exerted upon the assumption (shared as well by Schopenhauer) that immediately experienced guilt is *justified*.

of which we are responsible for our actions is unconscious (Žižek 2009, 188–190). But this stands in tension to Schelling's view as I read him, for on my reading, in guilt we feel both our freedom and our necessity, so that we *do* become in some sense aware of our original choice, which implies that even though the original deed was possibly unconscious, it is not completely inaccessible to consciousness at present.

3

Schopenhauer

The Varieties of Guilt

Introduction

This chapter turns to examine Arthur Schopenhauer's views on guilt as well as his version of the transcendental argument found in Kant and Schelling. As we shall see, Schopenhauer by and large treads the path opened up by Kant and Schelling before him. I first provide an outline of Schopenhauer's metaphysical view and then turn, in the second part, to examine the manner in which he understands the intelligible/empirical character distinction. In these first two sections, I rely mostly on his *The World as Will and Representation*. Needless to say, the discussion of Schopenhauer's worldview in what follows is not meant to be exhaustive but will provide us with the materials needed to make sense of his thoughts about guilt, which I turn to examine in the chapter's third part. Here I will distinguish three conceptions of guilt in Schopenhauer and show with the aid of his *On the Basis of Morality* and *Prize Essay on the Freedom of the Will* that the conception of *personal* culpability is grounded for Schopenhauer by an eternal and free deed of self-constitution where an individual chooses her own intelligible being or character as *causa sui*.

I. Schopenhauer's Metaphysics in Outline

We can approach Schopenhauer's metaphysical picture by viewing it in light of Kant's. While not remaining absolutely

uncritical of Kant's philosophy, Schopenhauer claims that besides the Hindus and Plato, it was "the impression made by Kant's work" to which he owes the "best in [his] development" (WWR I:417). He further adds that his own "line of thought, different as its content is from the Kantian, is completely under its influence, and necessarily presupposes and starts from it" (WWR I:416–417).

As is well known, Kant, in his *Critique of Pure Reason*, tries to show that traditional metaphysical aspirations to transcend empirical knowledge and go beyond what can be learned on the basis of experience should be severely limited, since on close analysis of our cognitive capacities, it becomes apparent that by its very nature, our mind is incapable of acquiring such transcendent knowledge—knowledge, for example, of the inner essence of the world. We thus have no cognitive access to the world as it is independently of how our mind experiences it by means of our categories and forms of intuition: we can have no knowledge of the "thing in itself." It is exactly to this restriction placed on our knowledge by Kant that Schopenhauer reacts in his philosophy.

In the first book of the first volume of his central work, *The World as Will and Representation*, first published in 1819, Schopenhauer adheres to much of the Kantian theory of knowledge briefly outlined earlier: he argues that every experience of the world, that is, every representation, necessarily presupposes a representing subject who perceives the objects of experience by ordering them according to the a priori forms of space, time, and causality. These a priori forms Schopenhauer calls the *principium individuationis* (WWR I:113), that is, the principle of individuation. This is so since it is only by means of time and space that we could be presented in experience with distinct individual objects. For example, we can experience two chairs as two distinct individual things only by virtue of each occupying a different place in space. In addition, as in Kant, it is by virtue of causality especially

that the phenomenal realm for Schopenhauer is thoroughly deterministic and all events are necessary. As he puts it: "all that happens, be it great or small, happens with complete necessity" (WWR II:319).

Nevertheless, despite his Kantianism with respect to cognition of phenomena, and while completely aware of Kant's restrictions on metaphysical knowledge, Schopenhauer attempts to grasp the nature of the inner essence of being, to comprehend the thing in itself. Indeed, he takes the task of philosophy to be precisely to answer the "riddle of the world" (WWR I:427), that is, the riddle regarding the inner nature of reality beyond appearances. According to Schopenhauer, our representations tell us only half of the story about reality, and the less important one at that; it is the essence of things, what lies behind their appearances, that constitutes the meaning of being and is thus what ought to be uncovered by philosophers who seek a complete and thorough understanding of the world. And this is precisely what Schopenhauer seeks to unearth. And so he holds, for starters, that while the world appears to us as constituted by individual, discrete entities interacting causally in space and time, this is just an *illusion*, the veil of Maya (e.g., WWR I:8), while underneath the veil, there is no distinction and no separation between one entity and another, between ego and alter. In the thing in itself, in *real* reality, all is one.

But if the world as we experience it is necessarily organized by us and is thus cognitively colored by our subjectivity, how can we secure unmediated access to the way the world is as such, to the thing in itself? How can we fathom its nature? Schopenhauer's answer is: *our body*. As he claims in section 19 of WWR I, while the body of each of us is given to us in just the same way as any other object, namely, as a representation which is ordered by means of time, space, and causality, it nevertheless has a unique status in that we can have a totally different kind of access to it, one that enables us to know it not as a mere appearance but as a thing in itself: we

can have access to it *from within*. In other words, “the body occurs in consciousness in quite another way, *toto genere* different,” and it is exactly this “double knowledge of our own body which gives us information about the body itself . . . about what it is, not as representation, but as something over and above this, and hence what it is *in itself*.” And what the body is in itself, indeed, what the *world* is in itself, “is denoted by the word *will*” (WWR I:103).

What is the will for Schopenhauer? Here is a summary of the central themes. First of all, the will lies beyond the *principium individuationis*. This means that the will is not determined by the forms of representation. As such, it transcends space, time, and causality. It is one (it appears as a plurality only in the realm of space and time), and it is groundless, that is, not determined by causality (since, again, causality operates only in the realm of phenomena) (WWR I:112ff.). This also means that it is unmotivated, since to be motivated for Schopenhauer means to be caused to act by means of representations (e.g., WWR I:36–37). Further, the will, Schopenhauer holds, is *blind*, *incessant striving*. The will is *incessant*, for, since it makes up the essence of the world, for the will there is no respite: were the will to halt, the whole world would be turned to naught. The will is a form of *striving*, analogous to but not identical to human desiring (the latter is an expression of the former); it is a constant movement-toward, a constant and insatiable wanting. Moreover, as a thing in itself, it has no particular object that it strives to attain as a subject, for the subject-object distinction makes sense only within the phenomenal realm. Finally, the will is *blind*; it is not conscious of its own nature and has no awareness of what it is doing or why (WWR I:110–111). It is only with human consciousness and reason that the will can come to have an awareness of itself. But awareness of ourselves and the entire phenomenal world as an expression of this desirous and insatiable will is of such a terrible nature, Schopenhauer holds, that if experienced deeply and properly, it can lead to the will recoiling back from itself in horror, which would amount to a denial of the

“will to live” (WWR I section 68)—a denial of the will as manifested in organic nature.

What is it, exactly, however, that is given to us once we turn our attention inward and try to view our body from the “inside” rather than from the “outside”? After declaring in section 19 that the inner nature of our body is the will, and after claiming that the rest of the world of appearances, just like our body, is only the external shell of the same reality, Schopenhauer claims in section 20 that

The will . . . proclaims itself first of all in the voluntary movements of this body, in so far as these movements are nothing but the visibility of the individual acts of the will. These movements appear directly and simultaneously with those acts of the will; they are one and the same thing with them, and are distinguished from them only by the form of perceptibility into which they have passed, that is to say, in which they have become representations. (WWR I:106)

In other words, it is the voluntary movements of our body—the turning of the head, the raising of the arm—that provide the initial clue; in them we experience the willing that underlies all reality.

But further, Schopenhauer holds that it is also when we turn our gaze inward that we find the will in itself in its incessant striving manifesting itself to our inner eye in a wide array of ways:

desiring, striving, wishing, longing, yearning, hoping, loving, rejoicing, exulting, and the like, as well as the feeling of unwillingness or repugnance, detesting, fleeing, fearing, being angry, hating, mourning, suffering, in short, all affects and passions. (FW 10)

And it is because, as manifestations of the one underlying will, we constantly will in these various ways, that we constantly are in *pain* to some extent. As Schopenhauer explains: “The basis of all willing,

however, is need, lack, and hence pain” (WWR I:312). This is one of the central premises that ground his famous pessimism.

Yet the question remains. How is it possible to have access to the thing in itself by means of internal perception? Why think that all these affects and passions, all these “manifestations of willing,” secure our grasp of the will as a thing in itself, underlying all our instances of willing? After all, Kant made it very clear that even internal perception is organized by the form of intuition of *time* so that access to the thing in itself is blocked even when we turn our reflective gaze inward. The question of the basis of Schopenhauer’s purported access to the thing in itself and its identification with the will is one of the most vexed topics in Schopenhauer interpretation, and interpreters’ claims range from criticizing Schopenhauer for making the gross mistake of trying to know the unknowable and identify the thing in itself with the will¹ to attempts to save Schopenhauer’s position by the introduction of various distinctions he himself does not make in his text, at least not explicitly.²

What is remarkable is that Schopenhauer is perfectly aware of the problem, as the following quotation from the second volume of *The World as Will and Representation* makes abundantly clear:

Meanwhile it is to be carefully noted, and I have always kept it in mind, that even the inward observation we have of our own will still does not by any means furnish an exhaustive and adequate knowledge of the thing-in-itself. It would do so if it were a wholly unmediated observation . . . but this knowledge of the

¹ See Janaway 1989, chap. 7.

² See Young 1987. Young argues that Schopenhauer makes use of a three-tier picture according to which besides the completely unknowable thing in itself, on the one hand, and empirical representations, on the other, there is a level of interpretation of the world that is intermediate, one level “above” the thing-in-itself and one “lower” than that of appearances, and it is this intermediate level which Schopenhauer characterizes as the will. Robert Wicks (2008, chap. 6), following Schopenhauer (WWR II:196–197), claims that since the will is perceived in inner perception, and since inner perception, in contrast to external perception, involves only the form of time, it allows us to view the thing in itself through fewer “veils,” so to speak.

thing-in-itself is not wholly adequate. In the first place, such knowledge is tied to the form of the representation; it is perception or observation, and as such falls apart into subject and object. . . . But on this very account, this I is not *intimate* with itself through and through, does not shine through so to speak, but is opaque, and therefore remains a riddle to itself. Hence even in inner knowledge there still occurs a difference between the being-in-itself of its object and the observation or perception of this object in the knowing subject. But the inner knowledge is free from two forms belonging to outer knowledge, the form of *space* and the form of *causality* which brings about all sense-perception. On the other hand, there still remains the form of *time*, as well as that of being known and of knowing in general. Accordingly, in this inner knowledge the thing-in-itself has indeed to a great extent cast off its veils, but still does not appear quite naked. (WWR II:196–197)

This answer remains far from satisfactory, of course, for it does not explain why the form of time is not as opaque as to obstruct *any* access to the thing in itself. Does this mean we are left with an argument that is ultimately based on a mysterious intuition of our essence?

The chapter “On Man’s Need for Metaphysics” in the second volume of *The World as Will and Representation* provides a possible answer. Here Schopenhauer says that despite Kant’s restrictions, “there are still other paths to metaphysics” (WWR II:182). What are these other paths? Schopenhauer says that “the whole of experience is like a cryptograph, and philosophy is like the deciphering of it, and the correctness of this is confirmed by the continuity and connection that appear everywhere” (II:182). Further, the “true interpretation of [the] meaning and content” of experience is “the metaphysical, in other words, that which is merely clothed in the phenomenon and veiled in its forms, that which is related to the phenomenon as the thought or idea is to the words”

(II:183–184). Schopenhauer's view is therefore that it is possible for us to gain access to the inner nature of reality, to the thing in itself, by interpreting the phenomena as if they were "a document the script of which is unknown" (II:184). By closely studying the appearances, it is possible for us to get at their inner nature, which expresses itself in the phenomena, and thus answer the riddle of the world. "The discovered answer to a riddle," Schopenhauer further explains, "[would] show itself as the right one by the fact that all the statements of the riddle are consistent with it" (II:184–185).³

Schopenhauer's argument for the identification of the thing in itself with the will thus seems to be *an inference to the best explanation*. He argues that while our knowledge of phenomena is inevitably shaped by our subjective cognitive capacities, what is given to us in knowledge is not exhausted by our subjective contribution. To deny the latter would be to fall prey to a thoroughgoing idealism à la George Berkeley. Therefore, the phenomena are always at the same time expressions of that which lies beyond our ken, namely, of the thing in itself. Consequently, "it must be possible to interpret these from it, and indeed from the material, not from the mere form, of experience" (II:183). And the correct interpretation would be the one that best fits all the data and illuminates it.

How, however, does this argument relate to Schopenhauer's initial move, namely, the discovery of the will in our body? I want to suggest that for Schopenhauer, it is in the feeling that we have of our body that we discover the first and crucial clue—a metaphysical Rosetta Stone, so to speak—to the deciphering of the code that is the world of appearances. This clue we can then use as a key to unlock the meaning of all other phenomena and consequently answer the riddle of the world. The view from within is only a first step, a springboard from which we can set off on our metaphysical quest.

³ Schopenhauer makes similar claims in the first volume in his examination of Kant's philosophy (WWR I:427–428). He can be read as offering a kind of fallibilistic metaphysics which, as he says, is not "conclusive" and complete.

In other words, I wish to suggest that Schopenhauer only helps himself to an “intellectual intuition” of sorts in order to get his discussion off the ground, but ultimately, the question of whether Schopenhauer is justified in identifying the thing in itself with the will depends on whether the analyses he provides of a wide range of phenomena—natural (organic and inorganic), aesthetic, moral, religious—are satisfactory, persuasive, and enlightening. At the end of the day, the cogency of Schopenhauer’s argument in favor of the will does not depend on a single argument he makes or fails to make; the entire *World as Will and Representation* (as well as Schopenhauer’s other writings) is one long argument in favor of the thesis and as such could not be settled at a glance but requires a detailed engagement.⁴

II. Empirical and Intelligible Character

Before I turn to discuss Schopenhauer’s conception of guilt and his version of the transcendental argument, I take a look at his understanding of the Kantian empirical/intelligible character distinction. According to Schopenhauer’s reading, this distinction lies at the very basis of Kant’s thinking about the compatibility of necessity and freedom, and he declares: “I entirely subscribe to this view” (FW 73).⁵ As we shall see, however, Schopenhauer reads the notion of the intelligible character in the stronger, Schellingian manner, namely, as concerning one’s very being (rather than one’s fundamental maxims). Schopenhauer comments on Schelling’s relation to Kant in his *Prize Essay on the Freedom of the Will* (published together with *On the Basis of Morality* in 1841):

⁴ Importantly, Schopenhauer argues (WWR II:183) that even this form of argument could only give us an answer that is partial and tentative; this is so since even here, our access to the thing in itself is constrained by the data given to us.

⁵ Even though he is critical of Kant’s manner of posing the problem about the compatibility of necessity and freedom in the latter’s third antinomy. See FW 21–22.

In his *Philosophische Untersuchungen über das Wesen der menschlichen Freiheit* Schelling furnished an explanatory paraphrase of Kant's highly important doctrine regarding the intelligible and empirical characters. . . . By the vividness of its color, this paraphrase may make the matter clearer to many than Kant's thorough but dry exposition. But I cannot mention it without reproving Schelling out of respect for truth and Kant. For he states one of the most important, most admirable, and, in my opinion, most profound of all the Kantian doctrines, but he does not say clearly and expressly that what he is expounding belongs in terms of its content to Kant. (FW 74)⁶

There are a couple of peculiarities that should be noted with respect to what Schopenhauer writes here. First, in light of our discussion in chapter 2, we can see that it is not quite accurate to hold with Schopenhauer that Schelling merely gave a "paraphrase" of Kant's view and was merely restating Kant's "wisdom" (BM 111); as we saw, Schelling's solution to the coexistence of freedom and necessity is quite radical and involves, first, a stronger metaphysical reading of a person's intelligible being and, second, a synthesis of freedom and necessity in the act of self-constitution of this intelligible being. Second, as noted and as we shall see in more detail, the manner in which Schopenhauer himself conceives of the intelligible character is in itself quite Schellingian and to that extent similarly goes beyond the Kantian view. Third, Schopenhauer berates Schelling for not acknowledging up front that he adopts the intelligible/empirical

⁶ In BM, Schopenhauer repeats these claims: "I regard Kant's doctrine of the coexistence of freedom and necessity [which Schopenhauer ties closely to the intelligible/empirical distinction; see WWR I:289] as the greatest of all the achievements of the human mind. With the Transcendental Aesthetic it forms the two great diamonds in Kant's crown of fame, which will never become extinct. It is well known that Schelling, in his treatise on freedom, gave a paraphrase of that doctrine of Kant, which by its vivid coloring and graphic descriptions is for many more comprehensible. I should commend this work if Schelling had had the honesty to say that he was stating Kant's wisdom and not his own. Even to this day a section of the philosophical public considers it to be his own idea" (BM 111).

distinction from Kant, but as we saw, Schelling openly claims that he is indebted to Kant when he presents himself as “expressing . . . the Kantian concept not exactly in his very words, but indeed in the way, as we believe, that it would have to be expressed to be comprehensible” (Schelling, *Phil. Inv.* 49).⁷ I will return to Schopenhauer’s critical comments about Schelling.

I turn now to Schopenhauer’s understanding and use of the empirical/intelligible distinction. As we saw, Schopenhauer holds that every voluntary action is an expression of one’s will. A more complete statement of the view is that every act of will that is manifested by bodily motion is initiated by *motives*, where a motive is a representation of an inner or outer state of affairs or a general thought “that stirs and moves my will” this way or that (BM 99) “through the medium of the intellect” (110; see also FW 28ff.). Motives can thus explain a person’s action: a motive *causes* the performance of an action with *necessity*, given that both motive and action belong to the realm of phenomena. This, however, does not provide a full explanation of the action. Schopenhauer claims (e.g., FW 41–42) that while a psychological explanation can track the different motives that causally bring about specific acts of will that manifest themselves in bodily action, there will always remain an extra principle intransigent to scientific explanations which explains *why* we will in the specific ways we do given specific motives. In other words, an action could be traced back causally to certain motives, that is, to causally efficacious representations of certain states of affairs, but this does not sufficiently explain why those motivations caused *this* specific action rather than another. Why does one person hand a bag of cash found on the street to the police while another keeps it to himself or herself? The answer to this question, Schopenhauer argues, cannot be found in yet another representation that the

⁷ A more substantial criticism Schopenhauer directs at Schelling concerns the latter’s “detailed account of a God with whom the worthy author betrays an intimate acquaintance. . . . It is only to be regretted that he does not say a word on how he arrived at such an acquaintance” (FW 75).

agent has, for two agents could be identical with regard to their representations and still act differently.

The answer, Schopenhauer holds, lies in the agent's *character*, a "particularly and individually determined quality of the will" (FW 42) by virtue of which a given motive *necessarily* drives a particular agent to will in the particular manner that he or she does. According to Schopenhauer, each person possesses such an *inborn, unalterable, and unique* character which constitutes his or her own identity as the specific individual he or she is (e.g., FW 42–48). Moreover, since this character becomes known only through experience and not a priori (FW 42), it is *empirical*: it is "only through experience" that "we come to know [our empirical character], not merely in others but also in ourselves" (FW 42). Specifically, each such individual character manifests itself as a set of spatial-temporal bodily phenomena, namely, as the set of actions which a person performs as a result of the causal effect of the motivations that he or she is affected by. The empirical character thus serves as a kind of individual psychological law that governs an agent's action and can thus ground explanations and predictions. Schopenhauer thus claims that every "deed of a human being [is] the necessary product of his *character* and the *motive* that has entered. If these two are given, the deed inevitably ensues" (FW 50). And as it is possible to predict, given a law of nature and the starting position of a system, the state of the system at a later point in time, so it is possible, Schopenhauer holds, given sufficient acquaintance with the character of a person, to predict how he or she will act in light of a certain motive. Thus, as Schopenhauer was fond of putting it, *operari sequitur esse*—our actions follow from what we are. This closely follows Kant's conception of the empirical character.

However, since the empirical character is empirical, that is, given to us in space and time, it must be itself the phenomenal expression of an underlying reality that lies *outside* of space and time. It must be itself an expression of something that is inborn, unalterable, and

unique and yet is not given to us through experience—this is the *intelligible character*: “For our grasp and comprehension, this character is *empirical*, but for this very reason it is only *phenomenon*; on the other hand, whatever it may be according to its real nature in itself is called the *intelligible character*” (BM 112). This intelligible character Schopenhauer also calls “the essence in itself” (BM 112), where the essence or *esse* of a thing is its peculiar nature: “For everything that is must have a nature essential and peculiar to it, by virtue of which it is what it is, which it always maintains and the manifestations of which are called forth by causes with necessity. . . . [A]ll this is just as true of the human being” (FW 51). Accordingly, the empirical character is “the mode and manner in which the very essence of our own self exhibits itself to the faculty of cognition” (FW 86–87). As Schopenhauer further claims, the empirical character is “determined by the intelligible” (BM 110). For Schopenhauer, then, actions arise by necessity out of the workings of a motive and a person’s empirical character, which is in turn determined by the intelligible character.

Difficulties with the empirical/intelligible distinction appear when we try to make room for the intelligible character within Schopenhauer’s metaphysics. A first problem arises precisely with respect to this relation of determination that obtains between the intelligible and the empirical; this is the familiar Kantian problem of how an intelligible, transcendental cause can affect the phenomenal realm. Second, it is hard to find metaphysical room for the intelligible character in Schopenhauer’s system: on the one hand, the intelligible, by definition, lies outside of space and time, but on the other hand, it cannot be simply identified with the one will that is the essence of all being, given that the latter is utterly undifferentiated and thus cannot serve to distinguish one essence from another, or one *individual’s* essence from another’s, as is required by the concept of character and the explanatory work it is supposed to do. Schopenhauer expresses his view about the will and character, for example, here:

man also has his *character*, from which the motives call forth his actions with necessity. In this way of acting his empirical character reveals itself, but in this again is revealed his intelligible character, i.e., the will in itself, of which he is the determined phenomenon. (WWR I:287)

It is clear from this that the empirical character is the intelligible one revealed in space and time, but how does the latter stand with respect to the will as such? In the quotation just given, Schopenhauer seems to virtually identify the intelligible character with the will, but for the reason given earlier, this just cannot work. In fact, this ambiguity in Schopenhauer—whether the inner essence of the individual human being is to be identified with the will in general or whether this will in general should be distinguished from a particular manifestation of the will, that is, the person's intelligible character—will be encountered again in the discussion of guilt.

As a solution to the problem of how to think of the relation between the will, the intelligible, and the empirical, we can hold that the one, general, undifferentiated will, which is the essence of being, expresses itself as a person's intelligible character—which indeed Schopenhauer occasionally refers to as a *person's* individual will (as when he uses an expression such as “my will,” WWR I:107)—which in turn is revealed, phenomenally, in one's empirical character. The intelligible character thus occupies an intermediate metaphysical position between the deep noumenal and the superficial phenomenal. A problem with this suggestion, however, is that as an *individual* being, a person's intelligible character cannot possibly lie outside of the phenomenal field of space and time, for individuation can by definition only manifest itself *within* space and time, the *principium individuationis*.

A possible way to think about this issue is to draw on Schopenhauer's notion of the *idea*, which he thinks of in Platonic fashion as an “objective image” (WWR II:364) lying beyond space and time particulars, and as containing, so to speak, the general

essence or the “true content” (WWR I:184) of its various individual phenomena. While the idea, for Schopenhauer, “does not reveal the being-in-itself of things” (WWR II:364), that is, the will, it is nevertheless an “immediate and adequate objectivity of the thing-in-itself, of the will” (WWR I:184). Thus, for example, there is the idea of the general essence of mankind (see WWR I:224ff.), but there is also—Schopenhauer claims—an idea of an individual person: “[It] is one of the distinguishing features of mankind that therein the character of the species and that of the individual are separated so that . . . each person exhibits to a certain extent an Idea that is wholly characteristic of him” (WWR I:224). And a bit later: “the human individual as such has, to a certain extent, the dignity of an Idea of his own” (225).

The idea, in other words, occupies a similar intermediate position between the empirical and the noumenal: like the intelligible character, it is not, on the one hand, in space and time but expresses, on the other hand, the underlying thing in itself, the will. Moreover, the idea, like the intelligible character, is a unity that “falls into plurality by virtue of the temporal and spatial form of our intuitive apprehension” (WWR I:234). Further, as Schopenhauer claims, an individual person has, “to a certain extent,” an idea of his own, which implies that there are also individualized ideas in Schopenhauer’s system. It is then perhaps possible to think of the intelligible character as a Schopenhauerian idea. The problem with this suggestion is that it does not help with our initial worry: how can ideas, as singular entities, be individuated independently of space and time? Schopenhauer, in my view, does not provide a satisfying response to this problem.

Be that as it may, given Schopenhauer’s view regarding the strict determinism that governs our actions, where whatever we do is necessitated by character and motives, the question arises whether and in what way a person is ever morally responsible for her actions and whether, consequently, anyone should ever feel guilt or bad conscience for her actions. After all, it seems there

could not possibly be any room for responsibility in a world where everybody's actions are determined with absolute necessity by their innate and unalterable character and the motives working on them: given the situation in which a person finds herself, no person can act differently from how she does, and therefore should arguably not be considered morally responsible and should not feel guilty. How, then, could the man born wicked, for example, be held responsible for his wicked deeds? Why should he feel guilty for those deeds? And, moreover, why *would* he feel guilty (in case he does)? How to explain the sting of conscience?

III. Three Kinds of Guilt

a. Repentance and the Spurious Conscience

To prepare for our discussion of Schopenhauer's view, we have to distinguish, on the one hand, what he calls "repentance" (*Reue*, also translatable as "remorse") from what he calls "pangs of conscience" (*Gewissensangst*), "sting of conscience" (*Gewissensbiß*), or guilt (*Schuld*). Let me start with repentance. As Schopenhauer explains, repentance over our actions occurs when we come to realize that as a result of false beliefs that acted as motives or some other cognitive disturbance, we "did something different from what was in accordance with [our] will," that is, he explains, when we come to realize that we have not acted "in accordance with [our] character" (WWR I:296). Here Schopenhauer seems to regard a person's inner being as a person's (intelligible) character, not as the will *simpliciter*. An example of repentance that Schopenhauer gives is of a person who acts more egoistically than in his true character because he was influenced by "exaggerated notions," was misled by others, failed to deliberate properly, or suffered from emotions excited in the present moment, with the result that he subsequently, upon obtaining a clearer, more truthful conception of the situation he was in, comes

to feel repentance over his egoistic action. Repentance thus does not arise “from the fact that the will has changed—this is impossible—but from a change of knowledge” (WWR I:296).

Schopenhauer’s view of repentance raises a difficult question: how is it possible to act *not* in accordance with one’s (intelligible) character given that, with the assistance of a motive, it determines one’s actions with a necessity of a law of nature? A possible solution, which seems to be faithful to Schopenhauer’s view (WWR I:297), is to claim that for Schopenhauer, one’s character truly comes to light when the agent’s beliefs about the situation he is in are true, when his deliberation is cool and sound, when his passions are restrained, and so on. Conversely, when an agent acts under one or more of these cognitively debilitating conditions, what is expressed is not his genuine character but a distorted projection, as if through a broken lens. This, however, cannot be of help, for by excluding the tendency to fail to deliberate properly or to be emotionally affected by present impressions, this suggestion threatens to dilute the very concept of character and leave it dry of content. For a person’s character seems, among other things, to include precisely such traits as the ability to be coolheaded, take things in the proper perspective, maintain one’s composure under pressure, and so on.

John Atwell makes a laudable yet ultimately unsuccessful effort, in my view, to solve this problem. He claims that when one acts, say, over-altruistically, the motive combines not with one’s particular, individual character but with “a human character (or with an instance of *the* human character) . . . in which I, as a human being, participate” (Atwell 1990, 65). In my view, this is not significantly different from, and thus not more intelligible than, explaining the unexpected behavior of a nitrogen atom, say, by referring to its participation in the nature of atoms in general. Further, why would the motives circumvent a person’s particular character and solely affect the human one? Perhaps, then, to make sense of Schopenhauer’s concept of repentance, one would need to limit it to cases of actions

that arise *only* out of false beliefs, that is, false motives. Actions based on false beliefs acquired not as a result of one's own fault indeed do not tend to manifest one's true character. The problem, however, is that this does not align with the text, where, as we saw, the circumstances for repentance are much broader.

Another problem with Schopenhauer's account of repentance is that it leaves us with no explanation of why, given his universal phenomenal determinism, one should feel repentance at all; after all, given the knowledge one possessed at the time and the manner in which this knowledge, as motive, affected one's actions, it was impossible at the time to have acted differently and in line with one's character. Importantly, Schopenhauer clearly believes that universal phenomenal determinism by itself undermines all moral responsibility (e.g., FW 49)⁸ and consequently guilt (FW 65). Why should one then feel the sorrow of repentance if one could not have acted differently? Lastly, it would make sense to feel repentance in Schopenhauer's sense only if one held it as an ideal to always act in line with one's intelligible character, to always be true to oneself or authentic. But Schopenhauer, it seems, does not tell us about the origin of this ideal, the grounds of its alleged authority over us, or its place in his metaphysical system.

Before I turn to Schopenhauer's discussion of guilt or pangs of conscience, I wish to briefly look at Schopenhauer's remarks about the conscience. First to be noted here—so as to distinguish Schopenhauer's view from Kant's—are Schopenhauer's critical comments about Kant's conception of the conscience, a conception I analyzed and criticized in chapter 1. Schopenhauer's central objections are very similar in nature to the ones I raised; thus, in *On the Basis of Morality*, he objects to the judicial metaphor Kant employs (BM 105), claims that the self-judgment Kant ascribes to the working of the moral conscience is not unique to the moral

⁸ Which is not to say that Schopenhauer denies all moral responsibility—the latter, however, is to have a different, non-phenomenal basis, as I will explain.

conscience but arises whenever a person ruminates painfully over what he or she has done (BM 106), and criticizes Kant for arguing from his description of the conscience to the necessary representation of a divine subject as a scrutinizer of hearts (BM 107–108).

Later in the same work, Schopenhauer makes his famously dismissive comments on the sting of conscience. In the section entitled “Skeptical View,” Schopenhauer discusses the question of whether it is possible that “there is no natural morality at all that is independent of human institutions” (BM 121). His conclusion is that in the overwhelming majority—though not the totality—of cases, “moral actions” are performed on egoistic grounds and are as such motivated by fear of God or the law, by concern for one’s social standing, and so on. He then goes on to consider the following objection to his rather cynical view of humanity:

In face of the *skeptical view*, appeal is made to *conscience*, but doubts are raised even against its natural origin. At any rate, there is also a *conscientiae spuria* [a spurious conscience] that is often confused with the genuine. The remorse and anxiety that are felt by many a man for what he has done are often basically nothing but the fear of what may happen to him. . . . Generally speaking, every inconsistency, every thoughtless action, every action contrary to our resolutions, principles, or convictions . . . every stupid blunder, afterward secretly annoys us, and leaves behind a sting in our hearts. Many a man would be astonished if he saw how his conscience, which seems to him such an imposing affair, is really made up. It probably consists of one-fifth fear of men, one-fifth fear of the gods, one-fifth prejudice, one-fifth vanity, and one-fifth habit. (BM 127)

Schopenhauer is in fact mixing different kinds of cases here, not all of which appear equally suspect (why does acting against one’s principles generate a spurious conscience?). But in light of Schopenhauer’s sharp words in the quotation, one might think

that for Schopenhauer, the sting of conscience as such is always spurious in this way or another and consists in a cocktail of various motivations that are alien to the one true motivation, the one that grounds *genuine* bad conscience or guilt, namely, the *moral* motivation. Schopenhauer, however, believes that such genuine, moral guilt, though rare, is possible and real. I turn now to examine Schopenhauer's account of such genuine guilt.

b. Schopenhauer's Three Forms of Guilt

In this section, I explain how in Schopenhauer's thought, we can find three distinctive yet interrelated kinds of guilt. Schopenhauer, however, does not clearly separate them from one another and occasionally runs them together in his discussions. As a first step, let's compare repentance to pangs of conscience or guilt: "Pangs of conscience over past deeds are anything but repentance; they are pain at the knowledge of one's own self in itself [*seiner selbst an sich*], in other words, as will. They rest precisely on the certainty that one always has the same will [*daß man den selben Willen noch immer hat*]" (WWR I:297; translation slightly modified).⁹ These words of Schopenhauer reflect the ambiguity encountered in the discussion of intelligible character. The problem is this: does Schopenhauer claim that guilt concerns knowledge of oneself as the one underlying thing in itself, the will, *or* does it concern knowledge of oneself as a *specific* manifestation of this will, that is, a certain *individual* intelligible character? Indeed, it seems that what is the same and unchanging cannot here be the one, undifferentiated will, given that as the thing in itself, it is the same in everyone and so cannot ground

⁹ Atwell formulates the difference thus: "repentance is pain over what we have done, and conscience [i.e., guilty conscience] is pain over what we are" (Atwell 1990, 234n70). Repentance, however, concerns only actions of a specific kind, namely, those that do not reflect our character, and bad conscience, on the other hand, concerns actions as well, namely, those reflective of our character.

guilt, given that the pains of guilt are not a constant throughout the lives of all human beings but depend on the particularities of what one does, which are in turn dependent, as we saw, on one's personal intelligible *character*. As I shall now explain, this ambiguity unfolds into Schopenhauer's three conceptions of guilt.

According to the first conception of guilt that we can find in Schopenhauer, we are all, as manifestations of the one insatiable will, *guilty as such*. This is Schopenhauer's philosophical interpretation of the Christian doctrine of original sin (e.g., WWR I:407, II:580, II:603ff.), his version of the idea of ontological guilt. The phenomenal world, Schopenhauer argues, is finite; the matter in which the will manifests itself is thus of limited supply. As such, the world's resources are limited; there is typically not enough of everything to go around for everyone. But since the will in its phenomenal objectification is desire for specific things, there is bound to be competition, conflict, strife, and pain. The multifarious forms the will takes battle among themselves for the satisfaction of their desires, only to renew the fight once a new wish appears on the scene. This is true not only of the animal kingdom, where besides the consumption of vegetables "every animal is the prey and food of some other" and "every animal can maintain its existence only by the incessant elimination of another's" (WWR I:147), but it is also the case in the human realm: humans, aside from subjecting nature in its entirety to their wishes, constantly fight against one another in the struggle to satisfy their needs and desires. In their attempt to preserve themselves, procreate, assert themselves, and attain wealth, fame, and standing, humans are at war among themselves, "and we get *homo homini lupus*" (I:147). The resultant picture, as Dale Jacquette memorably puts it, is of the world "as a monstrous coliseum in which every existent entity . . . tries, like a gladiator, to defeat every other existent entity" (Jacquette 2005, 115). The will, in its phenomenal embodiments, is thus at "variance with itself" (WWR I:146); "the will must live on itself, since nothing exists besides it, and it is a hungry will" (WWR I:154). In this way, merely

insofar as we are ourselves an objectification of the will, we also cause pain and suffering to others. Of course, it is possible that our individual lives will be such that the satisfaction of no desire of ours will rob someone else of the satisfaction of his or hers. But this is highly unlikely; in our search for a job, for prestige, for recognition, for a mate, for a place to live, and so on, we are bound to deprive someone of his or her desire and wish. As long as we affirm our will, as long as we desire, we cause pain and suffering to others (but also to ourselves, insofar as desire is constituted by pain for Schopenhauer). Being-in-the-world-with-others thus implies being hurt by others and hurting others. It follows, Schopenhauer thinks, that to will is to be guilty and thus that “existence involves guilt” (WWR II:569). Further, for Schopenhauer, the very fact that we suffer and ultimately die is in itself evidence of our guilt—it is the punishment existence inflicts upon us, so to speak, for existing. In Schopenhauer’s striking words: “If we wish to measure the degree of guilt with which our existence itself is burdened, let us look at the suffering connected with it. Every great pain, whether bodily or mental, states what we deserve; for it could not come to us if we did not deserve it” (WWR II:580). And: “the world itself is the tribunal of the world. If we could lay all the misery of the world in one pan of the scales, and all its guilt in the other, the pointer would certainly show them to be in equilibrium” (WWR I:352). This is Schopenhauer’s concept of *eternal justice*.

We saw that Kant endorses the idea that the human species is morally corrupt as such and that Schelling can perhaps be seen to depart from this view. Schopenhauer, in contrast, essentially agrees with Kant on this score, but for a different reason: he seems to think that whatever our particular characters are or whatever our fundamental maxim is, since we are all, in our particular wills, expressions of the one underlying thing in itself, we are all guilty precisely insofar as we exist and affirm this ravenous, relentless, and self-consuming will that is our noumenal nature.

We might wish to inquire why we *deserve* this punishment. After all, it seems, we did not choose to exist and participate in this horrendous show, so how can we be held responsible and judged guilty? Schopenhauer's response to this is that it is the will itself, existing out of space, time, and causality, that is free and "determines itself in itself and outside time" (WWR I:351). And "as the will is, so is the world. Only this world itself—no other—can bear the responsibility for its existence and its nature; for how could anyone else have assumed this responsibility?" (WWR I:351–352). In other words, it is the world that as noumenal will is free and therefore responsible for its own horrors which it manifests phenomenally. In addition, it is the world that judges itself to be guilty and deserving of punishment, and it is the world that also metes out the punishments in what Schopenhauer thinks is perfect proportion. The world is criminal, judge, and executioner all at once. This, however, does not address the problem. Again, why are *we*, as *individual* manifestations of this ravenous will, to blame? Why should we be punished? Schopenhauer can respond that this worry is misguided, since it assumes that we *are* individuals, while in truth, all individuality is a mirage; the *principium individuationis*, after all, is the cloth out of which the veil of Maya is made. But this response cannot be satisfactory, for even if all phenomenal individuality is a mirage, the pain and suffering aren't, for if they are, then the whole talk about the misery equaling the guilt would be nonsensical (illusory misery, being non-real, cannot equal anything). And if they are not, as seems obvious, then there must be real, distinct subjects that undergo this suffering, for the world as a representation as a *whole* cannot be considered a subject of suffering given that pain and suffering as such arise in the phenomenal realm out of particular desires of particular people for particular things. When discussing the third conception of guilt, I will return to this problem and present another reply that can be found in Schopenhauer.

A related issue that is more immediately relevant to my concerns here is that this conception of guilt as "original sin" does not account

for the first-person *experience* of guilt. After all, the latter, as empirical guilt, is experienced only occasionally and in light of specific actions of ours; it does not belong to us by virtue of our very being, and it does not haunt every instance of our lives.

Schopenhauer's second conception of guilt stands intermediately between the first and the third in that, on the one hand, like the first conception, it explains guilt mainly by reference to our deeper essence as the one general will, but unlike the first and like the third (which I will come to), it attempts to *explain* our empirical guilt, the painful feeling that arises (or can arise) as a result of our committing a specific action perceived by us as wrong in some way. The basic idea is that because distinctions between individuals are illusory, in all cases of immoral action, "tormentor and tormented are one" (WWR I:354), even though, at least initially, they might not be aware of this metaphysical unity: "The former is mistaken in thinking he does not share the torment, the latter in thinking he does not share the guilt" (WWR I:354). Let us focus on the former, the tormentor.

According to Schopenhauer, the perpetrator of evil deeds suffers from a peculiar kind of pain which is "associated with wickedness" (WWR I:364):

This pain is felt in the case of every bad action, whether it be mere injustice arising out of egoism, or pure wickedness; and according to the length of its duration it is called the *sting of conscience* [*Gewissensbiß*] or the *pangs of conscience* [*Gewissensangst*]. (WWR I:364–365)

Why, however, would the tormentor feel such pain? Schopenhauer distinguishes two sources for this pain. First, he holds that the tormentor

dimly sees that he, the bad person, is precisely the whole will; that in consequence he is not only the tormentor but also the

tormented, from whose suffering he is separated and kept free only by a delusive dream, whose form is space and time. But this dream vanishes, and he sees that in reality he must pay for the pleasure with the pain, and that all suffering which he knows only as possible actually concerns him as the will-to-live. (WWR I:365; see also WWR I:335)

The first source of the misery of the pangs of conscience arises out of the dim awareness that the pain of the other, caused by oneself as tormentor, is also one's own. It is the pain of being tormented by the tormentor that accounts on this view for the painful feeling of empirical guilt experienced by the tormentor himself—a feeling exacerbated the more the tormentor gains clearer metaphysical insight into the unity underlying the apparent diversity of I and thou. Empirical guilt, correctly seen, is thus a form of self-torment via the mediation of the other. Tightly related to this conception of guilt is Schopenhauer's view that it is the ability to see through the *principium individuationis* that stands at the basis of compassion (*Mitleid*), compassion being the only purely non-egoistic moral motivation, which consists in the desire not to harm another or to alleviate the other person's pain as if it were one's own (WWR I:374; cf. 387) because it is correctly seen to *be* one's own.

The second source for the pain of guilt is

associated with the first, namely, knowledge of the strength with which the will-to-live affirms itself in the wicked individual, extending as it does far beyond his individual phenomenon to the complete denial of the same will as it appears in individuals foreign to him. (WWR I:366)

The perpetrator recognizes the vehemence of his will, the tenacity with which he identifies with his will, and how it drives him to thwart the will of others who in essence are the same as himself. He thus experiences "an inward alarm at his own deed, which he

tries to conceal from himself [though] it contains that presentiment of the nothingness and mere delusiveness of the *principium individuationis*" (WWR I:366).

It is hard to say what we should make of this. My three main puzzles or objections are the following. First, this account seems radically remote from the first-person phenomenology of guilt: the experience of guilt for doing something wrong or morally wrong (the cases Schopenhauer has in mind here) does not apparently have as its content any insight into the unity of all beings or into the vehemence of one's will. Of course, it all hangs on the word "dimly" which Schopenhauer employs here to describe the tormentor's metaphysical insight (WWR I:365), for it allows him great argumentative freedom, but this freedom is directly proportional to the unpersuasiveness of the account. The second objection is that this conception of guilt does not clearly explain how the performance of an immoral deed brings about the metaphysical insight that Schopenhauer describes. What exactly is the psychological or metaphysical explanation for how the one brings about the other? Perhaps it is the apparent expression of the other's pain that stirs the tormentor's *compassion*, which, as we saw, leads the tormentor to see through the veil of illusion and thus feel the other's pain as his own—a feeling experienced as guilt. Guilt would then be almost reduced to compassion, and this looks like the wrong result. Third, this view of guilt does not clearly account for the feeling of *personal responsibility for the specific wrongful deed*. After all, all action as such is absolutely necessary for Schopenhauer and thus lacks all freedom. Similarly, a person's being a vehement manifestation of the noumenal will is not clearly his or her own personal fault, either (though see the following discussion of Schopenhauer's third view of guilt). And yet guilt seems to proclaim one's own personal responsibility for the deed performed. How is this to be explained?

It is in his third conception of guilt, I submit, that Schopenhauer makes the connection between empirical guilt, personal

responsibility, and freedom and provides us with his own version of the transcendental argument that we have identified in Kant and Schelling. Following his predecessors, Schopenhauer holds that while in the realm of phenomena all our actions—our *operari*—are strictly determined so that everything a person does is necessary, it is at the level of his or her intelligible character, or *esse*, that a person's "*true moral freedom*" (FW 83) is to be found. The evidence for this claim lies in the "feeling of *responsibility* for what we do, of *accountability* for our actions" (FW 83; see also BM 195).

How to explain this feeling of accountability, however, if all of our actions are necessarily determined? The only possible explanation, Schopenhauer holds, is that the feeling of responsibility, as a matter of fact, concerns not what we do but what we are—our character. Indeed, Schopenhauer holds, since "we are conscious of *freedom* only through the medium of *responsibility*, the former must also lie where the latter is to be found, and hence in the *esse* (what we are)" (BM 113). We are responsible for what we are, our character. But how are freedom and responsibility to be found in our *esse*, our character? Schopenhauer replies that it is in the thought that one "could have *been* a different man, and guilt or merit lies in what he is" (BM 112; cf. FW 84).¹⁰ It is thus here that we find the fundamental content of the feeling of empirical guilt. Schopenhauer explains:

It is true that the reproaches of conscience primarily and ostensibly concern what we *have done*, but really and ultimately what we *are*, for our deeds alone afford us conclusive evidence of what we are, since they are related to our character as the symptoms to

¹⁰ Wood describes Schopenhauer as holding that because we are fated to act as we do and could not have done otherwise, we are not morally responsible for what we do, "except in the sense that in the same circumstances, a different person, with a character different from [ours], would have acted differently" (Wood 1984, 94). This is only partly correct, in my view, for it does not emphasize the personal responsibility each person carries for *who he or she is*, according to Schopenhauer.

a disease. Thus guilt and merit must also lie in this *esse*, in what we *are*. (BM 195; see also BM 113)

Since our actions are all determined, responsibility and freedom have to do with the fact that we could have been a different person with a different (intelligible) character and consequently acted otherwise. We bear the responsibility for who we are as the specific individuals we are because we freely chose ourselves and are therefore guilty for this primordial free choice.¹¹ The guilt we experience over the wrong empirical actions we perform involves recognition of our responsibility for “what we are”—a responsibility that “presupposes a possibility of having acted otherwise and thus freedom in some way” (BM 110).¹² This original choice of who we are, we can then say, constitutes our own *individual* or *personal* ontological guilt, a guilt that marks our individual being and for which we are responsible.

But not only is empirical guilt an occasion upon which we mark our responsibility for what we are; it is in this empirical guilty that we also feel guilty *for* this fundamental act: “here in the *esse* (what he is) lies the point that is touched by the sting of conscience” (BM 113; see also FW 87–88, WWR II:604). In other words, empirical guilt concerns the specific immoral, phenomenal action that we perform only superficially; grasped more fundamentally, it pierces through the phenomenal screen

¹¹ I therefore disagree with Gardiner, who holds that Schopenhauer’s response to the question of whether we had any choice with respect to our intelligible character “could only be negative” (Gardiner 1963, 261). Gardiner seems to ignore Schopenhauer’s claims about moral freedom in BM, FW, and WWR.

¹² Schopenhauer, on my reading, has the conceptual tools to offer a *partial* response to Julius Bahnsen’s objection that since Schopenhauer “assigns individual differences to the realm of phenomena, and since he thinks that every action and event in that realm is determined according to natural laws, Schopenhauer cannot explain why we hold *individual* agents responsible for their action” (Beiser 2016, 241). On my reading, individual responsibility lies in one’s individual choice of oneself. This, of course, does not completely address the objection, since, as we saw earlier, the intelligible character is supposed to be individually differentiated as well, and how this is supposed to be possible, given that, as Bahnsen notes, individuation is phenomenal, is not clear.

and points to our choice of individual intelligible character because of which (with the help of specific motives) we performed the immoral deed. It therefore serves as a reminder of sorts of our choice of our individual *esse*: “Although the guilt lies in conduct, in the *operari*, yet the root of the guilt lies in our *essentia et existentia*, for the *operari* necessarily proceeds from these, as I have explained in the essay *On the Freedom of the Will*” (WWR II:604).

It thus seems that, like Schelling, Schopenhauer bases his claims about (moral) freedom on the acutely felt experience of guilt: our feelings of guilt whereby we become sharply aware of our responsibility and hence our freedom support the claim that we all acted freely on the intelligible level in the choice of our character, for it is truly there that our responsibility and freedom lie. This is Schopenhauer’s version of the transcendental argument. Here we see how, like Schelling, Schopenhauer bases his argument for freedom not on imputable original sin but on our accountability given to us in feeling—a felt accountability his argument presupposes as genuine and valid. Thus, even though throughout his *Prize Essay on the Freedom of the Will*, Schopenhauer’s answer to the question posed by the Norwegian Royal Society, “Can the freedom of the human will be demonstrated from self-consciousness?” (FW 3), seems to be no, at the end of the day, his answer is a clear yes: freedom, that is, intelligible freedom, can be inferred from our feeling of moral responsibility sharply experienced in the consciousness of guilt. As he puts it, “we *are* conscious of *freedom*” (BM 113; first emphasis added).

In addition, like Schelling, Schopenhauer holds that our empirical guilt ultimately indicates a deeper guilt that concerns a free act or deed whereby we chose to be what we are, our intelligible being. In other words, Schopenhauer, like Schelling, subscribes to the idea that at the intelligible level, we are our own free creation—we are *causa sui*. Here is another telling quotation:

It is easy to see that this path leads to our having to look for the work of our *freedom* no longer in our individual actions . . . but in the whole being and essence (*existentia et essentia*) of the human being himself, *which must be conceived as his free act*. (FW 87; final emphasis added)

To sum it up, Schopenhauer's version of the transcendental argument could be presented as follows:

1. We are subjectively aware of a feeling of personal responsibility, most prominent in the feeling of guilt;
2. Responsibility implies freedom;
3. Freedom is ruled out from the phenomenal realm, hence
4. Freedom is to be found in the noumenal or intelligible realm.
5. Specifically, to secure personal guilt and responsibility, this freedom has to do with the choice of one's personal essence;
6. That is, freedom is in the free choice of one's personal intelligible character, one's free choice of oneself, one's being *causa sui*.
7. It is in this eternal choice that our personal, ontological guilt lies.

Before I address a couple of objections to my reading, I wish to return to the first approach to guilt examined earlier: guilt as "original sin." There I asked for a justification for this, our ontological guilt, a guilt acquired by virtue of our very existence as embodiments of the will—a justification acutely needed given that, as it seems, we did not choose to exist. In the second volume of *The World as Will and Representation* (but this is possibly also indicated in the FW quotation from page 87), Schopenhauer addresses this worry directly by boldly claiming that we are guilty of this original sin because, in fact, we actually freely chose to exist! As he puts it, "According to the truth, the very origin of man himself is the act of his free will, and is accordingly identical with the Fall, and therefore the original sin, of which all the

others are the result, appeared already with man's *essentia* and *existential*" (WWR II:604; see also II:603). It is therefore possible to unite the third account with the first in the following way: every man is guilty *both* for his very existence (*existential*) as the kind of thing he is (*essentia*), namely, as an expression of the will, *and* for the choice of his particular intelligible character, the particular objectification of the will that he is, his individual *essentia*. For the former guilt, every person pays with the pain, suffering, and death that are the lot of every living thing; for the latter, one pays with the misery of the guilty conscience, which arises on the occasion of the performance of immoral deeds. For Schopenhauer, then, in contrast to Schelling and Kant, our original sin comes apart from our choice of intelligible being and constitutes a separate, deeper level of ontological guilt. In other words, on Schopenhauer's view, we could not have chosen differently and in such a way that would *not* burden us with at least *some* guilt: to choose existence is to burden oneself with guilt *period*, regardless of which character one chooses, an evil or a benevolent one. To completely rid ourselves of *all* guilt, we must deny the will to live—the only possible salvation.¹³

Let me now address a couple of objections. First, it might be argued that for Schopenhauer, the freedom required for individual moral responsibility and empirical guilt does not necessarily require for its establishment the datum of the feeling of bad conscience, for it follows from his basic metaphysical thesis regarding the will as a thing in itself: if the will is the noumenal and thus lies outside of space and time, it also lies outside the

¹³ Notice, however, Schopenhauer's claim that this denial of the will, like an "effect of grace," "is for us the only direct expression of the *freedom of the will*. It appears only when the will, after arriving at the knowledge of its own inner nature, obtains from this a *quieter*, and is thus removed from the effect of *motives* which lies in the province of a different kind of knowledge" (WWR I:404). The claim that this denial is the *only* direct expression of free will sits uncomfortably with the claim that the "very origin of man himself" is an "act of this free will" (WWR II:604), as well as with the view elaborated earlier that man is free in choosing his character.

realm of causality and is, therefore, free. In response to this, it must be emphasized that the freedom of the will as a thing in itself cannot secure *personal* responsibility and thus *empirical, individual guilt* for one's individual transgressions. Indeed, Schopenhauer distinguishes freedom *from* the forms of appearance (space, time, causality) and genuine *moral freedom*. Thus, he says that "the *thing-in-itself* underlying the phenomenon is outside space and time and free *from* all succession and plurality of acts" (BM 110; second emphasis added), while "*true moral freedom*" announces itself in a "fact of consciousness," namely, "the *feeling* of responsibility" (FW 83)—a feeling we most sharply experience in empirical guilt. The former is impersonal and has to do with the will being free from all determination, which in itself is not robust enough to ground individual moral responsibility for specific immoral deeds given that it amounts to mere blind, arbitrary striving. The second, however, involves one's own personal accountability—the choice of one's own individual essence.¹⁴

Second, some interpreters have found the claim that Schopenhauer locates responsibility and freedom at the level of our personal being to be philosophically indefensible and thus refused to attribute the view to him. Thus, we find Atwell arguing the following:

Consider what it would be like, within Schopenhauer's philosophy, for me to choose my own character. Since I am basically my character or will, if "I" were to choose the character or will that I am, "I" would have to antedate what I am, hence "I" would have to exist prior to existing—and this is self-contradictory. (Atwell 1990, 47).

¹⁴ I thus disagree with Janaway, for he seems to ground our personal intelligible freedom in the freedom of the will from space and time (Janaway 1989, 243–244), which undermines our personal responsibility.

There is no question that the idea of a person choosing himself or herself—the idea of *causa sui* at work, as we saw, in Schelling's argument as well—is philosophically puzzling, to say the least. There are, however, considerations that go beyond those adduced already that strongly support the interpretation that Schopenhauer indeed believed in the free choice of one's character, one's own intelligible being.

First, as if in response to the charge of philosophical absurdity, Schopenhauer openly admits at the very end of his discussion of freedom in his *Prize Essay on the Freedom of the Will* that, quoting Nicolas Malebranche, *la liberté est un mystère*—freedom is a mystery. And this is precisely because, “moved up from the domain of individual actions . . . into a higher region,” transcendental freedom “is not so easily accessible to our cognition” (FW 88). In other words, we should not be surprised if this freedom in the choice of who we are appears perplexing and even absurd. Second, Schopenhauer himself states—in a passage that is moreover quoted by Atwell himself—that his philosophy “is the only one that grants to morality its complete and entire rights; for only if the true nature of man is his own *will*, consequently only if he is, *in the strictest sense, his own work*, are his deeds actually entirely his and attributable to him” (WW II:589, emphasis in last two phrases added). A stronger and more explicit endorsement of the view could hardly be imaginable. Morality—the entire apparatus of praise and blame, guilt and responsibility—would not make any sense unless the human being is posited as *causa sui*. Third, immediately following his discussion in section 10 of his *On the Basis of Morality*, where he explains that freedom is to be found in our being, Schopenhauer adds a “Note” where he claims that “the Kantian doctrine of the intelligible and empirical characters” (which, as we know, Schopenhauer endorses) “is an insight, here raised to abstract clearness, which Plato had” (BM 113). Schopenhauer then proceeds to provide the reader with Porphyry's interpretation of the tenth book of Plato's *Republic*, where Porphyry's elucidation can be

seen, according to Schopenhauer, to be shedding light on the affinity between Kant and Plato on this point. The translation of section 39 from Porphyry's *Eclogues*, included in Schopenhauer's text, reads as follows:

For all that Plato means seems to be as follows. Souls, before they enter bodies and the different forms of life, have freedom of will to choose one form of life or another. . . . That freedom of will, however, is abolished as soon as the soul has slipped into any one of such life forms. For after the souls have passed into bodies, and from free souls have become living beings, they have only that freedom peculiar to the nature of the living being in question. (BM 115)

These sentences from Porphyry match Schopenhauer's view extremely closely: we freely choose our being, our intelligible character, but once chosen, our lives are determined, and we can only act in accordance with it.

As a final piece of evidence, I wish to return to Schopenhauer's criticism of Schelling's treatise on freedom. As we saw, Schopenhauer berates Schelling for presenting Kant's ideas about the empirical and intelligible character as his own. Schopenhauer goes on to comment that this has resulted in a predicament where "most readers, not intimately acquainted with the content of [Kant's] detailed and difficult works, are bound to think that here they are reading Schelling's own thoughts" (FW 74). Schopenhauer then gives as an example of this misattribution the words of the German philosopher Johann Eduard Erdmann, whom Schopenhauer quotes as writing that "Leibniz, like Schelling in his treatise on freedom, represents *the soul as determining itself prior to all time*" (FW 74; emphasis added). Schopenhauer then immediately adds that we can see how "Schelling here stands to Kant in the fortunate relation of Amerigo to Columbus; another man's discovery is labeled with this name" (FW 74). It is thus plain to see

that according to Schopenhauer, both Schelling and Kant adhered to the view that the soul determines itself in an eternal act, and since Schopenhauer himself explicitly proclaims full allegiance to the Kantian doctrine (FW 73), it follows that he, too, adheres to the same view. I conclude that Schopenhauer did indeed subscribe to the *causa sui* view.

But on what basis do we choose one intelligible character rather than another? Like Schelling, Schopenhauer believes we choose our being. But unlike Schelling, Schopenhauer does not state explicitly that we are necessitated by our essence to choose (freely) the essence that we are. Instead of attempting to delve into the ground of transcendental freedom, it seems Schopenhauer is willing to stop his reflections at the point of our free choice of our character and go no further. This, however, raises a problem: can Schopenhauer block the infinite regress that plagued Kant's account? Differently asked, how can Schopenhauer explain why we choose the intelligible character that we end up choosing rather than another character? Without helping himself to the Schellingian solution, without claiming that our free choice of who we are is necessitated by who we are, Schopenhauer, it seems, would be unable to block the regress without introducing arbitrariness (thus eliminating imputability). In this regard, Schelling's conceptualization of the human being as *causa sui* addresses the Kantian problematics in a more explicit and fuller manner.

And yet, if one is indeed cause of oneself, then, one might wonder, how is it possible for one *not* to be necessitated to choose as one does and create oneself differently? When one is a cause of oneself, there is no possible distinction between creator and created that can possibly allow for a divergence between the two and thus make it metaphysically possible for the creator to choose someone other than himself or herself. It thus seems that even though, unlike Schelling, Schopenhauer does not explicitly develop his conception of the act of self-creation, the very notion of *causa sui*

includes within itself the response to the Kantian problematics. Consequently, the idea of freedom that Schopenhauer employs to secure responsibility, like Schelling's, is not that of leeway freedom (the ability to choose differently).

IV. Conclusion

It is time to summarize the results of the first three chapters. The transcendental argument I have extracted from all three philosophers—Kant, Schelling, and Schopenhauer—presents an attempt to argue on the basis of our guilt, whether empirical or ontological, for a free deed that each person has performed. All three agree that determinism is incompatible with freedom, and so they all try to locate this free deed at the noumenal or intelligible level, where our fundamental or ontological guilt also lies—a guilt that justifies empirical guilt but can also serve to explain it. Despite this broad similarity, however, it is only Schelling and Schopenhauer who think of this original deed as involving an act of self-creation in a strong, metaphysical sense. Further, as we saw, Schopenhauer, unlike Schelling, does not explicitly conceive of this deed as grounded necessarily in who one is. It thus seems that despite Schopenhauer's identification with Kant's view and despite his charge that Schelling in essence merely reiterated Kant's position, all three reach similar though not identical views.

At the end of the day, however, all three versions of the transcendental argument for freedom that I examined in the first three chapters, while intriguing and philosophically rich, and despite the manner in which they admirably pursue the logic of guilt to its conclusion, fail to make clear sense of their posited noumenal act of freedom, which thus remains puzzling and mysterious. Another point in which they are all united is that they all uncritically assume that guilt in some sense is justified or imputable. In the next

chapters, when I turn to look at what I call the “naturalist tradition,” some of those central issues will be revisited and explored from a non-metaphysical standpoint that, in contrast to the “metaphysical tradition,” is extremely skeptical of the justifiability of guilt and freedom, as well as of the notion of *causa sui*.

4

Rée

The Naturalization of Guilt

Introduction

In this chapter, I consider Paul Rée's views on guilt,¹ responsibility, free will, and punishment. Three subsidiary aims of my discussion are to show that, first, Rée's works are worth paying attention to in their own right and not merely as a foil for Nietzsche's ideas; second and relatedly, that Nietzsche's criticisms of Rée's views fall short of the mark; and third, that in contrast to the views of some interpreters, Rée's thought undergoes development in the course of his career.

The discussion in this chapter and chapter 5 on Nietzsche will present us with a radically different approach to guilt from the one encountered in the first three chapters. This approach spurns all metaphysical speculation and attempts to consider guilt as a *natural* phenomenon, by which I mean a phenomenon that does not require for its explanation an appeal to some noumenal or intelligible realm where agents perform timeless deeds but is to be explained as a thoroughly this-worldly *psychological* phenomenon.² In Rée's and Nietzsche's thought, we find one of the first sustained attempts in German philosophy to naturalize morality in general and the

¹ Throughout, following Rée, I will not distinguish guilt, feelings of guilt, remorse, pangs of conscience, or bad conscience, unless indicated.

² I do not enter here into the thorny question of how to define naturalism; I explain it here mostly *negatively*, i.e., with reference to what it does *not* involve or what it leaves out, namely, the noumenal or metaphysical.

phenomena of guilt in particular on the basis of the conviction that all metaphysical speculation is a naive, immature, and unscientific way of thinking. In this, both philosophers open a new path for philosophizing which diverges from and opposes the traditional metaphysical views considered in the previous chapters. One of the main thrusts of this “naturalistic tradition,” as I call it, is that at the end of the day, guilt cannot be justified, though it can be *explained*.

Rée is usually not treated in the history of philosophy as a philosopher of interest and importance in his own right and is usually discussed in the context of his relationship to Nietzsche, who tends to eclipse Rée and leave the impression—bolstered, no doubt, by Nietzsche himself—that Rée’s ideas could be either ignored or quickly done away with on the way to the more philosophically rich and insightful Nietzsche.³ Though I, too, will relate Nietzsche’s ideas in chapter 5 to Rée’s, I wish in the present chapter to devote close attention to Rée, for I believe that his ideas—written in clear and compelling prose—are of value and interest in themselves and merit the attention of the philosopher and historian of philosophy. Indeed, it is precisely by virtue of his clarity of expression and straightforwardness of philosophizing that I think his contribution to a naturalistic understanding of guilt in some respects surpasses Nietzsche’s.

Though, as noted, I believe Rée’s thought is of interest independently of Nietzsche’s, I agree that examining the two philosophers’ views in successive fashion, as I will do here, is justified. As is well known, Nietzsche and Rée developed their ideas in a form of cross-fertilization in the 1870s. Specifically, after having first met in Basel in the summer of 1873, the two spent a period of five months together at the Villa Rubinacci in Sorrento under the general care of Malwida von Meysenbug in 1876–1877,

³ Janaway (2007), though he devotes some pages to Rée, treats him at the end of the day as a foil for Nietzsche’s ideas, as, of course, it makes perfect sense in a book devoted to Nietzsche’s *On the Genealogy of Morals*.

during which time Rée composed his *The Origin of the Moral Sensations* (*Der Ursprung der Moralischen Empfindungen*, published in 1877), which he later dedicated to Nietzsche with the words “To the father of this work most gratefully from its mother.”⁴ Nietzsche reciprocated and exclaimed in a letter to Rée from August 10, 1878: “Long live Rééalism and my good friend!”⁵ It was also during that period that Nietzsche continued to work toward his *Human, All Too Human* (published in 1878), a text that seems to betray a heavy influence of Rée, in both style and content.⁶ And despite Nietzsche’s attempts to dissociate himself to some extent from Rée, both at the time of HH and later,⁷ it is clear that he also held Rée’s philosophy in great esteem and describes him (in HH 37) as “one of the boldest and coldest of thinkers” and, several years later, as “excellent” (EH, Books, *Human*, 6). Considering their views side by side will enable us to better appreciate each thinker’s position and contribution and the manner in which they differ from and complement each other’s ideas.

I start by examining Rée’s naturalized explanation of the experience of guilt as expounded in his *The Origin of the Moral Sensations*. I then turn, in the chapter’s second part, to look more closely at Rée’s critique of Charles Darwin on the origin of bad conscience. In the third part, I examine the development of Rée’s thinking in his later *The Emergence of Conscience* (*Die Entstehung des Gewissens*, published in 1885), and in the fourth part, I briefly discuss his *The Illusion of Free Will: Its Causes and Consequences* (*Die Illusion der Willensfreiheit: Ihre Ursachen and ihre Folgen*, published in 1885).⁸

⁴ See Small’s introduction to Rée 2003, xiii.

⁵ See Small’s introduction to Rée 2003, xxxviii.

⁶ See Small’s introduction to Rée 2003, xxxiv–xliii, for a valuable discussion, as well as D’Iorio 2016. Both cite letters where Nietzsche protests this claim of influence and argues that he had already started to compile notes toward HH even before he moved to the Villa Rubinacci with Rée.

⁷ See note 6 above and, e.g., GM P 4; EH, Books, *Human*, 6.

⁸ References to *The Emergence of Conscience* and *The Illusion of Free Will* are to the German edition of Rée’s *Gesammelte Werke* (Rée 2004).

I. Rée on the Origin of Pangs of Conscience in *The Origin of the Moral Sensations*

In the preface to *The Origin of the Moral Sensations*, Rée clarifies immediately that the purpose of his book is purely *explanatory*—a “purely theoretical one”—and that

[j]ust as the geologist begins by seeking out and describing different formations and then inquires into the causes from which they have arisen, so too the author has begun by taking up moral phenomena from experience, and has then gone into the history of their development [*Geschichte ihrer Entstehung*], as far as his abilities allowed. (*Origin* 85; translation slightly modified)

This statement of purpose already contains the germs of Rée’s (and to a great extent Nietzsche’s, as we shall see) naturalistic mode of philosophizing. This approach, for Rée, consists of two central components. First, the naturalist philosopher begins with a set of observations about his field of study, in this case “moral phenomena,” describing them as thoroughly as he can. Second, he turns to examine the causal history or the genesis of these phenomena, reconstructing it “as far as his abilities allow.” This historical narrative, to use a term of Nietzsche’s, amounts to a *genealogy* of the phenomena.

In the introduction to *Origin*, Rée claims that the moral phenomena he will focus on are the following: “that certain actions are felt as good, others as evil [*als gut, andere als böse empfunden werden*]; that evil actions often give rise to pangs of conscience [*Gewissensbisse*]; that on account of the so-called sense of justice [*Gerechtigkeitsgefühl*] we demand punishment for bad actions [*schlechte Handlungen*]” (*Origin* 87).⁹ He goes on to clarify that

⁹ Small chooses “bad” for the German *böse*, but I prefer to abide by what has become customary practice in Nietzsche scholarship and reserve “bad” for the German *schlecht*

while these phenomena are “often considered to be something supersensible,” “the voice of God,” or “transcendent” (87), he intends to put in place of such “transcendent explanations” ones based on “immanent causes,” an approach that is now made possible in light of the recently discovered theory of evolution. “[T]oday,” he adds, “since Lamarck and Darwin have written, moral phenomena can be traced back to natural causes just as much as physical phenomena: *moral man stands no closer to the intelligible world than physical man*” (87; emphasis added). This italicized judgment—later favorably quoted by Nietzsche (HH 37)—already indicates the vast philosophical distance that separates Rée (and Nietzsche) from the metaphysical approach which, as we saw, upheld the exact opposite view. In contrast to metaphysical philosophizing, which seeks “transcendent” explanations in an “intelligible” realm, the basic approach Rée adopts in his attempt to uncover the origination of the various moral phenomena he lists is the immanent and empirical theory of evolution, which “rests essentially on the following proposition: The higher animals have developed by natural selection from lower ones, for instance, human beings from apes” (*Origin* 87). But though Rée will indeed erect his explanation on broadly Darwinian foundations, his full account of the phenomena is multilayered and involves not only a basic biological level but also a sociological, psychological, and even a metaphysical one, but where it is the *belief* in certain metaphysical claims (though not their justification or truth) that plays a crucial explanatory role. In addition, as we shall see, Rée is not a blind follower of Darwin but subjects the latter’s view on the nature of guilt to criticism.

I will now turn to reconstruct Rée’s account in *Origin* of the emergence of the experience of pangs of conscience/remorse (*Gewissensbiss*) or consciousness of guilt (*Schuldbewusstsein*)—Rée treats these as the same. I break down my reconstruction into the

and translate the German *böse* as “evil,” even though Rée himself does not distinguish “bad” from “evil” as Nietzsche does in his *On the Genealogy of Morals*.

following steps: (a) the evolutionary origin of the basic instinctive endowment: drives and feelings; (b) the utility-based valuation of the instinctive endowment: the origin of the value concepts “good” (*gut*) and “evil” (*böse*) in both the nonmoral and the moral sense; (c) the origin of the *intrinsic* moral valuation of good and evil; (d) the feelings that come to be associated with intrinsic moral valuations; (e) the role of responsibility and free will; (f) Rée’s possible response to Nietzsche’s criticism.

(a) Rée starts with the claim that in the human being, two central drives (*Triebe*) are operative: “the egoistic drive and the non-egoistic drive” (*Origin* 89), where the latter is weaker than the former, by which he seems to mean that in a motivational conflict between the two, it is the egoistic drive that will prevail over the non-egoistic one, which will have to remain unsatisfied, at least temporarily.¹⁰ The egoistic drive, Rée explains, is concerned with one’s own self-preservation, the satisfaction of one’s own sexual needs, and the satisfaction of one’s own vanity (89) and can thus motivate actions that are harmful to others, though not necessarily. The non-egoistic drive, on the other hand, takes the “welfare of others [as] the final ends” of action so that a person who is motivated by this drive either “seeks their welfare for their own sake or refrains from harming them for their own sake” (89). After arguing for the reality of actions motivated by the non-egoistic drive (89–91), Rée turns to consider an explanation for the possibility of such actions. He sketches Schopenhauer’s metaphysical account, where, as we saw briefly in chapter 3, to act non-egoistically is to act from the insight that self and other are metaphysically one, and argues that this account “must give way to Darwin’s simpler explanation” (92). It thus seems—though Rée does not elaborate on this—that considerations of simplicity at least partly drive Rée’s preference for the naturalistic, Darwinian view in general and

¹⁰ Rée also refers to these drives as feelings (*Empfinden*), desires or appetites (*Verlangen*), and motives (*Motiven*).

Darwin's explanation of the non-egoistic drive in particular, which Rée summarizes as follows:¹¹

Like many animal species, such as bees and ants, our ancestors the apes have a social drive. Brehm [Alfred Edmund Brehm, 1829–1884] says, “The mutual attachment of the members of a tribe (of chimpanzees) is very great. The males love the females, the females their children to a remarkable degree, and the stronger always defend the weaker.”¹² This social instinct arises from an extension of the parental instinct and is then maintained and strengthened by natural selection, that is, by the fact that the animal species whose members were most closely bound together by social instincts displaced other species and so continued alone. As human beings developed from the apes, they thus already had the drive to care for other human beings (at first, the same community). Someone who has this drive will feel pleasure at seeing other people happy (non-egoistic sympathetic pleasure) and pain at seeing them unhappy (non-egoistic compassion). (92)

Rée starts, then, with the assumption of a parental instinct in apes to care for their young, the origin of which is supposedly to be accounted for by the greater survival chances of those who are taken care of by their parents, which would then explain how this instinct is passed forward to the next generation. This, however, would only explain the origin of non-egoistic concern for one's descendants. Rée therefore claims (following Darwin)¹³ that this instinct extends beyond the family unit to encompass other members of the group, and this presumably—the text seems to imply—by means of the

¹¹ Interestingly, Rée does not stop to provide an analogous Darwinian explanation for the origin of the *egoistic* drive, perhaps because he believed that it raises no great explanatory mystery. He holds, however, that it is “older and stronger” than the non-egoistic drive, which is “the later and the weaker” (*Origin* 96).

¹² Rée is quoting Brehm 1864–1869, 1:25.

¹³ See Small's introduction to Rée 2003, xxvi.

mechanism that has come to be known as “group selection”: the behavior motivated by the non-egoistic instinct increases the strength of the group as a whole, the thriving of which, in turn, implies the greater survival chance of each of its members, thus better securing the transmission of the non-egoistic drive to future generations.¹⁴ Finally, this instinct, given that we have evolved from apes, gets passed down (or up) to us, human beings, and thus accounts for the existence of *our* non-egoistic drive—the drive to help others for their own sake.

But besides the tendency to behave in these non-egoistic ways, there is also, according to Rée, as the long quotation just given mentions, an *affective* component to such non-egoistic behavior. Specifically, he claims that action motivated by compassion involves the feeling of *pain*, the pain caused by the *image* or *representation* of the other’s suffering, as well as pain over the other’s suffering *itself*. He then goes on to add that

non-egoistic actions are also accompanied by feelings of *pleasure*. In fact, when we help a suffering person for his own sake, what disappears at the same time as his suffering is our painful mental picture of his suffering, and our pain over the fact that he is suffering; and there arises in us the pleasant feeling of having performed a good deed. . . . [I]f we had failed to perform the good deed, we would perhaps later have felt pangs of conscience [*Gewissensbisse*]. (90; emphasis added)

And Rée later adds that we also feel a similar pleasure to the one just mentioned when we *witness* the well-being of others: “someone

¹⁴ But then it is not clear why one needs to posit the paternal instinct at the origin of one’s explanation of the non-egoistic instinct. Why cannot one explain its origin directly by reference to group selection? Further, Rée, in the quotation given (92), seems to hold that the instinct first finds its expression in parental concern for one’s offspring and then “extends” to cause other-regarding behavior in general, a behavior that is *subsequently* “maintained and strengthened by natural selection.” No explanation is given, however, for how this initial “extension” occurs.

who has this [i.e., the non-egoistic] drive will feel pleasure at seeing other people happy (non-egoistic sympathy)” (92). Though I will put it aside for now, we should notice the important claim made at the end of the quotation just given: failure to act on one’s non-egoistic drive can occasion pangs of conscience at a later time.

In addition to these more naturally occurring affects, Rée also holds that a more specific kind of feeling attaches to our actions. Thus, actions of the egoistic kind “are felt by each of us as morally bad and blameworthy [*empfindet jeder von uns als moralisch schlechte und tadelnswerth*], whereas [non-egoistic actions] are felt as morally good and praiseworthy [*als moralisch gut und lobenswerth*]” (93).¹⁵ In other words, I want to suggest, Rée distinguishes feelings that arise naturally in us as a result of what we do or observe and feelings that are associated with *value concepts* of what is morally good/bad or praise-/blameworthy. Since the feeling of pain involved in the valuation of our actions will be crucial for our understanding of Rée’s account of the experience of guilt, it is important to ask about the origin of these concept-associated feelings. I will first turn, however, to consider Rée’s account of the origin of the concepts themselves—the concepts of the good/bad in both the nonmoral and the more developed moral sense and the related concepts of the blame-/praiseworthy. I return to the subject of feelings in section (d).

(b) Rée’s account of the origin of the valuation of certain actions as good or bad or as praiseworthy or reprehensible in a *nonmoral* sense is rather straightforward and is summarized by him with the following words:

¹⁵ It is important to emphasize that while words such as “reprehensible” or “blame-worthy” seem also to connote the idea that the agent of such actions *deserves* in some way to be reprimanded or blamed, and thus presuppose a notion of responsibility and desert, Rée for the most part, it seems, wishes to use them merely as meaning something like “bad.” The ideas of responsibility and desert he develops separately, as we shall see.

Egoistic actions that occur at the expense of others were originally condemned on account of their harmfulness; non-egoistic actions were originally praised on account of their utility [*Nutzens*]. Later, the former were condemned in their own right, and the latter in their own right. (*Origin* 99)

Thus, actions motivated by egoism were originally dubbed—conceptualized as—“bad” or “reprehensible” (in a nonmoral sense) because they harmed the well-being of particular others or of the community as a whole, while actions motivated by the non-egoistic drive were dubbed—conceptualized as—“good” or “praiseworthy” (in a nonmoral sense) on account of their usefulness to particular others or the community as a whole. However, as Rée makes clear, “it is not only non-egoistic behavior, but often self-interested behavior as well, that is useful to others: for example, a physician who heals for money may achieve as many cures as the physician who helps others out of selfless motives” (94), and yet this type of person or action, while good, is not called *morally* good; “rather . . . this description is applied only to those who are useful to others, or do not harm them, *out of benevolence*” (94; emphasis added). In other words—to underscore—what is morally good is not simply identified with the useful.

So the question now arises with respect to the emergence of this specific distinction between “good” in the sense of useful and “good” in the sense of *morally* good (and similarly for “bad”). This involves, Rée holds, taking into account the *motives* of actions. Rée claims that in the beginning of human history, no such distinction between the good and the morally good was made at all, and he cites John Lubbock’s *History of Civilization and the Primitive Condition of Man* (1870) in support of the claim that “savages, when they make any moral judgment at all, never consider the motives of actions, but only their usefulness or harmfulness” (*Origin* 96). It is only later,

at a higher stage in the evolution of knowledge [that] one must instinctively have felt that, if someone refrains from harming others only out of fear [of punishment], peace is imposed from outside and hence exists only on the surface. . . . On the other hand, when people refrain from harming others not out of fear but for their own sake, peace is not imposed artificially from the outside but comes from the inside. (96)

Rée is making here a subtle point it is important to linger on: he is arguing that while from the point of view of utility, both non-egoistic and (some) egoistic actions can be seen as useful and in that sense as good, those that are motivated solely or mostly by the non-egoistic drive come to acquire a higher valuation and come to be *morally* praised because it is perceived that the utility of actions motivated by egoism, though certainly not negligible, is “accidental and uncertain” (97). It is accidental and uncertain because the egoist does not care for the well-being of others as such but only regards it as a means to further his own interests; once he realizes that there are more expedient ways for him to get what he wants, he will cease to care for others. His actions’ utility is thus contingent on whether others’ well-being *happens* to serve his egoism. In contrast, the utility of actions motivated by non-egoism is “certain” and “necessary,” for here one’s desire to help others is not dependent on whether helping them happens to serve one’s needs; one helps others for their own sake. And Rée adds: “Indeed, if everyone had truly non-egoistic feelings, if everyone loved his neighbor as himself, then communism would not only be possible, it would already be present at hand” (97). Unfortunately, this is not the case; the egoistic drive is much more prevalent and stronger than the non-egoistic drive. But this does not stop people from comparing the way things currently are to a state where everyone loves their neighbor as themselves. People thus come to describe non-egoistic behavior as “desirable, praiseworthy—good” (97), now in a *moral* sense. Rée thus importantly clarifies that “non-egoistic behavior would never

have been held up as the good if what is at hand coincided with what is desirable" (97). In other words, Rée's point—though admittedly, he does not make it clearly or consistently—seems to be that "good" in the moral sense signifies not actions that are merely immediately useful (such utility is in principle provided by some egoistic action) but actions that are non-egoistically motivated and thus believed to help society progress in the long run toward its perfection. To call a benevolent action morally good is thus not only to commend the action itself for its immediate utility but also, and more importantly, to express the view that because of its non-egoistic motive, it is of higher value, for if this motive gained ascendancy in society and outweighed the egoistic drive in its vehemence and prevalence, society would be better off as a result of this motive's behavioral effects. The morally good "brings us closer to a state of greater happiness" (98).

As a consequence of the realization that it is non-egoistically motivated actions that tend to improve the overall condition of society, Rée claims (97), people start to inquire into the *motives* of actions and consider only those actions that are motivated (solely or mostly) by the non-egoistic drive to have *moral value*—to be *morally good*. On the other hand, those who harm others out of selfish motivations are considered *morally bad*, and those who are "useful to others out of egoism" are "not described as morally praiseworthy" (98), for while their actions might be good in the sense of useful to others, they are nevertheless motivated by an egoistic drive the expressions of which are perceived as not bringing society "closer to a state of greater happiness" but rather as tending to preserve its wretchedness.¹⁶ In this way, "the degree of non-egoistic feeling and behavior . . . became the criterion of moral value" (97).

¹⁶ At one point, Rée claims, implausibly in my view, that the person who refrained from doing harm to others out of egoistic motivations began at one stage to be regarded as *just as blameworthy* as the person who actually harmed others (*Origin* 97), since both were considered to be motivated by egoism and, more specifically, by the desire to harm others.

Alternatively put, we reach a stage where “*only* egoistic actions are called [morally] bad, *only* non-egoistic actions are called [morally] good” (93). Importantly, at this stage, the utilitarian perspective is still adhered to: non-egoistic actions are seen as morally good because they are not merely useful but are *more securely* useful, in contrast to egoistic actions which are either downright harmful or not reliably useful.

(c) In the next stage of his reconstruction, Rée explains how this utilitarian perspective is abandoned. He thus claims that “nowadays, however, we do not praise the good because of its useful consequences, but instead it appears to us praiseworthy in its own right, independently of all consequences” (*Origin* 98). Crucially, now, the morally good or bad is not conceived of in utilitarian terms but in quasi-Kantian ones: certain kinds of action are right or wrong intrinsically or as such. This development occurs, Rée explains, because while originally non-egoistic action was praised with an eye to its long-term utility, “afterward people became accustomed to praising it and forgot that this praise was originally based on usefulness to the community” (98). Non-egoistic actions thus become valued as good *in themselves*. Egoistically motivated actions, on the other hand, become valued as either lacking any moral value (when not harmful) or bad *in themselves* (when harmful). Now that this more developed distinction in moral value has come into being,

people set out to impress it upon children. . . . We constantly hear the selfless person praised and the egoist condemned. The books we read and the plays we see present the same opposition; finally we are directly taught that unselfishness, compassion, benevolence, and sacrifice are good, and that hard-heartedness, envy, and malicious pleasure are bad. (100)

As a consequence of this instruction and upbringing, we come to forge almost unbreakable associative links between the ideas of the *morally good* or *praiseworthy* and the *unegoistic*, on the one hand,

and the ideas of the *morally bad* or *blameworthy* and the *egoistic*, on the other. Rée explains that as a result of this, whoever has come to acquire such associations “involuntarily associates the satisfying *feeling* of having done what is good and praiseworthy with his own non-egoistic actions as well, and also associates the painful *feeling* of having done what is bad and blameworthy with his own egoistic actions” (102; emphasis added). In other words, it is now made evident that in associating non-egoistic actions with the morally good and egoistic actions with the morally bad, we at the same time come to associate a certain *feeling* with our conceptual valuations of such actions. Rée, however, does not really explain how this association of feelings develops.

(d) We can fill in the details on Rée’s behalf in the following way. The idea is that conceptual moral valuations attain a certain affective content in the process of upbringing. Specifically, when a child acts non-egoistically, adults will smile at her or express satisfaction in some other way and sometimes reward her in some more material way. Such responses will tend to make the child feel good, and she will come to associate such feelings with what are called by those same adults morally good, non-egoistic actions. In contrast, egoistic actions will be frowned upon by adults who will express anger and dissatisfaction with the child’s action and will sometimes punish her in some way. Such responses will tend to make the child feel bad, and she will come to associate such feelings with what are called morally bad, egoistic actions. As a result, actions of a certain kind will come not only to be thought of in a certain way but also to stir certain feelings in the child: a positive, satisfying, feeling will be felt upon the performance of moral actions and a negative, painful one upon the performance of immoral actions.¹⁷

¹⁷ One worry that arises here is that my attempt to help Rée and fill this gap in his explanation is doing more harm than good, for on my suggestion, the negative feeling associated with a “bad” action is impressed on the young mind of the child along with the (explicit or implicit) threat of punishment. Consequently, it could be that the negative feeling of guilt one comes to experience is not morally pure, for it is tinged with fear of punishment; but fear of punishment, Rée claims in his *The Emergence of Conscience*, as

Now Rée adds: “The more strongly someone feels [*fühlt*] that egoistic actions are bad, the more he must appear bad and blameworthy [*verwerflicher*] to himself when he has been incited by this egoism to commit such actions” (102). But “a man who appears bad and blameworthy to himself because he has inflicted suffering on another feels what is called pangs of conscience [*sogenannte Gewissensbisse*]” (102). In other words, the idea is that since I associate the concept of the morally reprehensible as well as a certain negative feeling with egoistic actions, when I realize that *I* acted egoistically (or when I failed to act non-egoistically as I should have), I consider my own action as *morally reprehensible* and *feel* a negative, painful feeling about it and consequently about *myself*—the feeling of guilt.¹⁸

Let me now pause to comment on Rée’s account. Even though, as we just saw, Rée moves swiftly from the claim that we feel negatively about our *action* to the claim that we feel bad about *ourselves* and thus experience guilt, this transition (i) is not self-explanatory and (ii) involves a specific and nontrivial conception of guilt, namely, as concerned ultimately with the person rather than with the action. With respect to (i), we can rightly inquire about the mechanism that is responsible for inflecting the negative affectivity from action to agent. One possibility is that this occurs by virtue of the agent holding the (Schopenhauerian) view that actions are expressive of

we shall see, is distinct from guilt. To fix this issue, we can hold, first, that in the course of time, the thought of punishment drops out of mind, while the negative feeling associated with the bad action remains; second, that an adult’s frown, say, can give rise to negative feelings in children that are not connected in their minds with any kind of punishment.

¹⁸ One thing to notice is that although Rée claims that guilt can also arise when we don’t act non-egoistically as we should have (90), an explicit account of why this should be the case is missing. Specifically, he does not take care to explain why a negative moral evaluation would attach not only to egoistic commissions but also to non-egoistic omissions. Presumably, Rée’s story, to accommodate this, would have to include an additional element: it was realized by the earliest generations of mankind that the *non*performance of non-egoistic actions also had the long-term effect of preventing society from approximating a state of social perfection. Consequently, such omissions were dubbed “bad” in the moral sense and came to be imbued with a certain negative feeling.

who one is. One realizes what one does, infers that this is indicative of who one is, and thus comes to feel bad about oneself. This, of course, in itself, is not a self-evident principle, and the fact, if it is a fact, that agents prone to guilt subscribe to it must be explained. With respect to (ii), there is disagreement about the proper object of guilt, with several writers¹⁹ holding the view that guilt does not concern the agent but is more focused on the action. Rée, unfortunately, does not address these issues. Nor is he consistent in his adherence to this view: as we shall see, in certain instances, for example, in his discussion of free will, Rée makes the agent's *action* the focus of her guilt.

(e) Rée's account of bad conscience or consciousness of guilt is, however, not yet complete. The full picture requires the ingredient of *responsibility*. Rée opens his discussion of responsibility in the third chapter of *Origin* with the following words: "Remorse [*Gewissensbisse*] differs according to whether whoever feels it bears in mind the necessary character of human actions or not" (104). He immediately goes on to list approvingly all the thinkers who have denied free will and concludes that the "topic can be regarded as settled" (104): free will is nonexistent. Nevertheless, he goes on to add, we tell ourselves, after having committed a misdeed, "I could have acted differently," and "so arises the deceptive illusion that commonly misleads people" (105)—the illusion of free will. It is this belief in free will, understood here by Rée as the belief that one could have acted differently, that is, the idea of leeway freedom, that forms the crucial ingredient in Rée's concept of responsibility. To regard oneself as responsible is to think that one could have acted differently at the time of action.²⁰ Thus, freedom (ability to act differently) is a necessary condition

¹⁹ Taylor 1985; Williams 1993b; Prinz 2012.

²⁰ Rée does not *define* responsibility in this way, but this could be inferred from his various claims to the effect that given that no one is free and could have acted differently, no one is responsible, and so no one should be held responsible for his or her actions (see *Origin* 111–112).

of responsibility, which in turn implies that one can rightly be held responsible for one's misdeed not only by others but by one-self as well, in the feeling of guilt.

But how exactly does this idea that one could have acted differently emerge? Rée explains (105–106) that because we believe we possessed at the time of action a *general ability or a capacity* to do the right thing, to reach the right conclusion, or to overcome our egoistic passions, we think that this ability or capacity could have also been effective at the time of our wrong action and could have enabled us to act differently. But this is wrong, for under the conditions existent at the time of action, our abilities or capacities could not but have been as they in fact were, namely, ineffective; to be effective, other causal forces had to be either present, so as to stir these capacities into action, or absent, so as to let them operate as they normally do, “since every effect must in fact have its cause” (106). Rée then claims:

Even though every act of will is necessary, knowledge of this fact is encountered only in very few people, that is, solely in those who know how to think. All the rest take their will to be free, and so their remorse [*Gewissensbisse*] ordinarily takes the following form. Someone reflects on an action he has committed . . . and associates the idea of blameworthiness [*Tadels und der Verwerflichkeit*] with this action. He then charges himself with this action [*rechnet er diese Handlung sich selbst zu*], in that, without reflecting that it resulted with necessity from the thoughts, sensations, and circumstances present at the moment when it was committed, he thinks unhappily, “I could have acted differently.” (107)

He then adds, “Remorse [*Gewissensbisse*] ordinarily attaches to a single action and retains its sting *because* we unthinkingly assume the freedom of the will. In most cases, then, remorse is based on error” (107; emphasis added).

The view Rée articulates seems to be this: when one erroneously believes that one could have acted differently at the time of action, one succumbs to the illusion of free will. But this means that one conceives of oneself as responsible for the misdeed and “charges himself with this action,” with an action, moreover, that one considers reprehensible (i.e., morally bad) and feels bad about. This mental complex—considering yourself responsible for an act you think is wrong and feel bad about—is the experience of pangs of conscience. In other words, the view presented in *Origin* is that the illusion of free will is a necessary condition of the experience or consciousness of guilt.²¹ Consequently, because of the falsity of the belief in free will, *guilt is unjustified*. Of course, as we saw in chapter 3, the metaphysical philosopher would have agreed that there is no free will in the realm of phenomenon and that *therefore*, for guilt to be justified (as it is for this philosopher), freedom on the intelligible level must be posited. For Rée’s naturalistic perspective, this is not an option: there is only this world where there is no freedom, and so consequently guilt is a mistake.

Interestingly, however, Rée thinks that something *akin* to pangs of conscience remains even for the person who has acquired philosophical insight and abandoned the belief in free will. Instead of thinking that he could have acted differently, this person understands that his actions necessarily resulted from the causes operative at the time of action, but “he remarks that the most essential [cause] is the constitution of his own *character*; and now he feels horror [*Entsetzen*] at the fact that he has a character from which the actions which he cannot help feeling as blameworthy could proceed” (108; emphasis added). In this case, Rée explains, the person’s quasi-remorse is attached “not to the particular act but to its origin in his character, not to the *operari* but to the *esse*” (108). Rée is clearly influenced here by the Schopenhauerian view where the experience of empirical guilt really concerns one’s character, attained freely in

²¹ As we shall see, in his later writings, Rée drops this view.

an intelligible deed (according to Schopenhauer). Rée, however, quickly proceeds to distance himself from Schopenhauer's view and explains that this horror is not based on some "mysterious intelligible freedom" and that we "have received our innate character not through any fault of our own" (108). We are merely horrified at the fact that we possess such a character that is able to produce morally bad actions. In other words, there is no sense of responsibility for who we are that lies at the basis of this horror.

Next, Rée considers a person who not only has grasped the truth about the unfreedom of the will but has also come to grasp an important truth about the origin of our valuations and has come to realize that the actions we denounce are neither "condemnable" nor "blameworthy" in themselves but "actions of a particular nature with which people originally associated the labels 'blameworthy' and 'condemnable' simply because such actions harm the community" (109). This person, Rée claims, will feel remorse only weakly (*nur schwach empfinden*); he will not be able to rid himself of it completely given the tight psychological link that obtains in his mind between such actions, their evaluations, and the attendant feelings.

Finally, Rée moves to consider the person who "has not at all become accustomed to consider as blameworthy those actions that harm the welfare of others"; this person "will not experience any remorse after committing such actions . . . but will experience at most a feeling of frustration" (109). But this feeling, "stripped [*entblösst*] of the sensations [*Empfindungen*] of responsibility, blameworthiness, and self-condemnation [*Selbsttadels*], is nothing but the sort of unsatisfied feeling that arises from the frustration of any drive, and so it cannot be described as remorse" (109). This, however, does not apply to most people living in "higher culture." For the latter, then, guilt would be an error not merely because of its presupposition of a free will but also because of the belief informing guilt that certain actions are reprehensible as such. This completes Rée's explanation of the origin and nature of remorse.

It is important to note that with the formulation given here, a new element appears on the scene, that of self-condemnation. We not only feel bad about ourselves when considering our action as morally wrong, but we also feel the negative feeling of self-*condemnation*. Whence this feeling of condemnation, however? Is it essentially, in Rée's view, the negative feeling we feel about ourselves when we consider ourselves as bad, or is it the latter but intermixed with the idea of holding ourselves responsible for the action (under the assumption of free will)? Rée does not explain what exactly he has in mind here. A further question concerns the feeling of frustration we just saw Rée mention. Why does *frustration* remain once all the moral conceptualization is peeled off the experience of guilt? I will return to this in the next part of this chapter when I examine Rée's criticism of Darwin's view of remorse.

(f) Before I turn to examine Rée's criticism of Darwin, I first wish to consider one criticism *Nietzsche* makes of Rée's account in his *On the Genealogy of Morals*—a criticism well familiar to readers of Nietzsche. In the opening sections of the first essay of the *Genealogy*, Nietzsche criticizes Rée's view without, however, mentioning Rée by name, referring to him obliquely as a member of the "English Psychologists," though the German Rée was perhaps only English by philosophical temperament.²² Nietzsche's main critical thrust is that the English psychologists'/Rée's historical account of the origin of the concept "good" does not make psychological sense. Rée argues, according to Nietzsche, that

Originally . . . one approved unegoistic actions and called them good from the point of view of those to whom they were done, that is to say, those to whom they were *useful*; later one *forgot* how this approval originated and, simply because unegoistic actions

²² See Janaway 2007, 78–80, for an argument for the view that Rée is the *singular* referent of the expression "English Psychologists."

were always *habitually* praised as good, one also felt them to be good—as if they were something good in themselves. (GM I:2)²³

Nietzsche, however, believes that such an explanation cannot work because it is hard to see how those who benefit from non-egoistic actions will forget why such actions have received the appellation “good,” given that, being their beneficiaries, they are constantly reminded of the positive utility of such actions (GM I:3). Nietzsche is right that such forgetfulness would be quite inexplicable. He is therefore right that it would be strange for people to forget why they started to call such beneficial actions “good.” But Nietzsche is misrepresenting Rée’s genealogy and simplifying matters.

First, as we saw, Rée does not hold that “originally” one approved of the unegoistic. Rather, originally, one approved of those actions that were deemed useful. It is only *later* that the evaluative perspective turned to focus on the action’s motivation. As Rée claims: “the distinction between the non-egoist as good and the egoist as bad is . . . not as old as the human race, but is the product of a later cultural stage” (*Origin* 141). Second, Rée’s view, as we saw, is that the forgetfulness he assumes applies not to the concept of “good” *simpliciter* but to the concept of the *morally* “good,” where the latter does not simply attach to the non-egoistic because of its immediately felt positive consequences. The immediate beneficial consequences are not the reason the non-egoistic is called morally “good,” since such consequences by themselves are not enough to distinguish the *morally* good from the good *simpliciter* that the egoistic can generate as well. Rather, as we saw, non-egoistic actions

²³ In his later *The Emergence of Conscience*, Rée can be seen as espousing a radical relativistic view that makes it clear that it is not exactly correct to attribute to him the view that it is the non-egoistic actions that come to be called “good” and the egoistic ones “bad”; citing anthropological evidence, Rée repeats again and again the claim that among the so-called lower civilizations, certain actions that we moderns condemn—*egoistic* actions—were originally praised (e.g., *Emergence* 219–226). In this regard, too, Nietzsche, I think, was being unfair to Rée’s more mature view, familiar to Nietzsche at the time of writing GM.

are considered morally “good” because they are seen as beneficial in a more permanent and reliable manner—they are more useful in the long run. But the long run is not something that is “an everyday experience at all times” (GM I:3). It is the extent to which the non-egoistic brings humanity closer to a “state of greater happiness” (*Origin* 98) that grounds the moral valuation “good,” which then, as a result of forgetfulness, gives rise to the view that the non-egoistic is morally good in itself. I thus believe that Nietzsche’s criticism misfires.

II. Rée’s Criticism of Darwin

As we saw, to complete Rée’s account of guilt, we had to provide him with an explanation of how certain conceptual valuations come to be endowed with certain (positive and negative) feelings. Without this affective element, it is not clear how Rée can explain the *sting* of conscience. In the course of laying out his theory of the emergence of guilt, Rée critically engages with and rejects Darwin’s view of the nature of remorse and his explanation of the negative feeling that lies at its core. According to Darwin, on Rée’s interpretation, when the non-egoistic drive is left unsatisfied and one acts egoistically instead, at a later time, upon recollection of the harmful effects of one’s actions, the agent’s non-egoistic drive awakens and “assert[s] its frustration and manifest[s] itself as a dissatisfied, painful feeling, as remorse” (*Origin* 102). Now, Rée agrees that “such a feeling of frustration must certainly be felt by people when the development of their intellect has enabled them to recall what has occurred in the past whenever they have failed to take account of their non-egoistic instinct” (102). But he quickly adds that

the feeling of frustration we experience when an instinct—in this case, that of non-egoism—is not satisfied is not remorse, any more than the feeling of frustration we feel when an egoistic

instinct is not satisfied, for example, if we have let slip an opportunity to satisfy our vengefulness. (102)

In other words, Rée holds, there is no essential difference in feeling between the frustration of our non-egoistic drive and the frustration of our egoistic one, which means that an account, such as Darwin's on Rée's reading, that aims to base remorse on the former fails to capture its unique nature which involves not mere frustration but "the feeling of blame and self-condemnation" (102). In other words, to account properly for remorse, one has to explain the origin of the concepts of the morally good and bad, the feelings that attend them, and the notion of moral responsibility, as Rée does.²⁴

Robin Small holds that Rée's objection to Darwin's account of remorse targets only a simplified version of his view. Darwin's more considered explanation of remorse involves "the unusual stability and constancy of the social instinct, its reinforcement by a desire for the approval of others, the communication of their expectations through language, a more or less extended process of upbringing and habituation, not only memory but foresight, and so on."²⁵ But it seems that even if we agree to add all the additional components that Small lists to the basic phenomenon of instinct frustration, we would still be several paces short of the finish line. One central problem is that the elements Small lists do not clearly help us understand in concrete terms how the feeling becomes morally conceptualized.²⁶

²⁴ An interesting implication of this criticism of Darwin's view is that non-concept-wielding animals cannot possibly feel guilt.

²⁵ Small's introduction to Rée 2003, xxix. Small also wonders what the pain of remorse could be based on according to Rée if *not* on the frustration of the moral instinct/drive (xxix). This is an extremely apt question, and I tried to answer it in the account I provided on Rée's behalf of the feelings that attend actions conceptualized as good/bad. As we shall see, in his later writings, Rée introduces the additional idea that remorse involves the thought that one's actions deserve punishment.

²⁶ In his book on Rée and Nietzsche, Small seems to express agreement with this point: "If Rée makes an advance on Darwin's account of the moral sense, it is by elaborating the process through which the social instincts innate in many species take a

Further problems beset the Darwinian account. Let's assume that the painful feeling of the pangs of conscience or guilt involves at its core the frustration of the non-egoistic drive, which, after the wrong deed has been performed, comes with the aid of recollection to haunt the egoistic person and make its protestations. Let's even further assume that the subject has at his disposal the concepts of "bad" or "reprehensible" and believes in his free will and thus in his responsibility for his actions. Thus, our "remorseful" subject now not only recollects that he failed to do the non-egoistic thing and comes as a result to feel the pain of this instinct's frustration, but he also comes to view the commission of the egoistic action as bad and as something for which he is moreover responsible. But notice that on this enriched Darwinian view, there is a disconnect between the painful feeling and the immoral act performed, between the presumed feeling of guilt and the fault; thus, one feels pain not *for having done wrong* but for *not* having done right, that is, for not satisfying the non-egoistic instinct of compassion. Certainly, guilt can arise over omission just as over commission, but it is precisely this important difference that is ironed out by the present account, where *every* instance of guilt feeling is a result of the former, and where the latter is, in a manner, guilt-free. The murderer would then be struck by guilt feelings not because he has taken the life of another but because his compassion did not find its satisfaction in saving the man's life—a compassion aroused by his (the murderer's) own doing!

We can perhaps address this problem by replying on behalf of the Darwinian account that we tend to misinterpret our inner experiences and fall into the illusion of thinking that we feel pain for having done wrong—in contrast to pain for not having done right—as a result of some process of psychological

association: since the act committed is more present to our mind than the act omitted, we think of the pain that in actual fact arises as a result of the latter (omission) as arising out of the former (commission). This solution, however, won't save the Darwinian account from other problems.

The Darwinian assumption is that when we feel remorse, this is because our non-egoistic instinct was frustrated, that is, left unsatisfied at the time of action when we instead chose to act egoistically and harm someone. But it is far from being obvious that all instances of bad conscience have this etiology. Specifically, the non-egoistic drive, the drive to help others or alleviate their suffering, typically arises only in specific situations, when we witness someone else's torment, for example. Put differently, when we act egoistically, the non-egoistic drive typically does not arise at all and thus does not suffer frustration of any sort. The feeling of remorse that is felt subsequently cannot therefore be explained with reference to the frustration of the non-egoistic instinct. Situations in which despite the arousal of the non-egoistic drive, we nevertheless act egoistically, thus leaving our non-egoistic drive frustrated, are situations typically characterized by a form of inner struggle, where we feel torn between the good and the bad and end up failing to do what is right. But, again, such a predicament fails as a general account of the emergence of all cases of remorse.

An alternative and perhaps more promising way to defend the Darwinian view is the following, suggested by Rée himself when he claims that "a man who allowed his children to die of hunger—assuming he had never learned that such an action is blameworthy—would have only a feeling of frustration *afterward*, a compassion coming too late, but not the feelings of blame and condemnation that characterize remorse" (*Origin* 103; emphasis added). While on the previous reading of Darwin's suggestion, the frustration of the non-egoistic drive occurred already in the past, when the agent failed to act compassionately and instead

inflicted harm on others, on the second view considered at present, the non-egoistic drive demands satisfaction *afterward*, not because its demands were previously silenced—the agent’s compassion was not stirred at all during the time of action—but because memory brings to mind images of suffering or pain that trigger the drive’s awakening in the present, causing it to appear and demand its satisfaction vociferously. Since, however, it cannot be satisfied *now*, for what’s done is done, the resulting feeling is that of the pain of frustration. We strongly desire to help the suffering people who appear in our recollections but cannot, for the time of action is past.

This account, too, suffers from the incongruence of experience and explanation mentioned earlier, for here, too, we would not feel remorse over what we have done. But this account presents us with an additional incongruity, for on the present view, we suffer remorse not over what we have done, and not even over what we have not done, but as a result of our *not-being-able-to-do-now*. Put differently, our pain, a result of our presently aroused compassion, is frustration over not being able to act *now* for those who suffered *then*. We imagine them now and picture their needy state now, and this arouses our compassion—and yet we cannot do anything for them. A more significant problem with the account presently under consideration is that our personal connection to the past action provided by the thought of personal responsibility—an indispensable ingredient in the feeling of remorse—would be severed: memory would serve us with reminders not of what *we* did or did not do but of something that happened and that we now cannot do anything about. The resulting feeling would be closer to what R. J. Wallace (2017) has termed “impersonal regret,” where something we value is damaged or hurt in some way, though not as a result of our own doing, and we feel pain for it having been affected in this deleterious way.

Even though Rée is right, in my view, to think that the Darwinian account cannot work, Rée, of course, agrees that

acting compassionately, or failing to so act—even independently of the moralization of the instincts—can result in certain kinds of feelings. As we saw, Rée clearly believes that pleasant feelings can arise as a result of performing a non-egoistic deed. In addition, he also thinks that the frustration Darwin had in mind could be experienced by people with a developed intellect. This seems to entail that, even though, as I argued, the feeling of remorse cannot be understood as the feeling of the frustration of the non-egoistic instincts, it might nevertheless occasionally be *accompanied* by such a feeling of frustration. Is this intermingling a phenomenon Rée considers psychologically possible or real?

Let me now conclude this discussion. Rée's account in *Origin*, while an important step in the direction of the naturalization of the feeling of guilt and a clear advancement beyond Darwin's view, is lacking in certain important respects that form the desiderata for any future account. First, as we saw, the negative feeling associated with the feeling of guilt is not given adequate explanation by Rée, and we had to fill this gap on his behalf. Second and relatedly, for Rée, as we saw, remorse involves "the feeling of blame and self-condemnation [*das Gefühl des Tadels und der Verwerflichkeit des eigenen Selbst*]" (102). But how can we explain this ability to turn one's emotions at *oneself*, and who or what is this self that one accuses, and what is its relation to the misdeed? Third, while Rée attempts to provide an explanation for how we come to the view that we possess free will, it remains unsatisfactory, for it fails to account for the error that people fall into when they think they could have acted differently: why do people believe that their faculties could have sprung into action at the right moment (even though, as a matter of fact, they did not)? Whence the source of this illusion? I will consider Nietzsche's views on some of these questions in chapter 5. I now turn to consider Rée's later analysis of the pangs of conscience.

III. The Emergence of Conscience

Rée's later book, *The Emergence of Conscience* from 1885, is considered by scholars as not significantly going beyond the material presented in *Origin*.²⁷ I believe this assessment is only partly fair, for there are nevertheless reasons to consult this later text as well in our attempt to understand Rée's view and its development. The most central aspects with respect to which this book differs from the earlier *Origin* are the following. First, *Emergence* includes much more historical and anthropological data in support of its claims—it is much more empirically grounded than the slimmer *Origin*. On the basis of this data, Rée can be seen to argue for a far more moral relativistic view than the one put forward in *Origin*. Second, even though Rée already theorizes the concept of punishment in *Origin*, in *Emergence*, he employs the concept of deserved punishment to enrich his conceptualization of guilt. Third, Rée introduces in *Emergence* certain distinctions that further expand his understanding of the phenomenon of guilt. Fourth, and crucially, Rée *breaks* the connection between the consciousness of guilt and the freedom of the will ascertained in *Origin* and almost completely omits any mention of free will from his analysis. While *Emergence* aims to explain both the phenomenon of pangs of conscience and the phenomenon of "good conscience," I will focus here on the former.

²⁷ Thus, Small is of the view that it does not contribute anything of substantial significance to the views developed in the older book (Small's introduction to Rée 2003, xlvii–xlviii). Janaway only somewhat concurs with this assessment, for he recognizes the introduction of the idea of punishment into Rée's account of the guilty conscience (Janaway 2007, 84). See also Nietzsche's scathing view of the book: "how empty, how boring, how false! One should write only about things of which one has one's own experience." Nietzsche then qualifies this judgment a bit and praises its "simple, clear, and almost antique form" and immediately denigrates it again by complaining that "it is a pity there is not more 'content' in such attire!" (letter from October 15, 1885, quoted in Small's introduction to Rée 2003, xl). One wonders whether Nietzsche himself adhered to the principle he himself lays down here and only wrote about things of which he had personal experience!

To begin, Rée provides us with a description of the phenomenon to be explained. He first claims that “Establishing an exact definition of conscience would be a mistake, since it would be more accurate than what people think under conscience” (*Emergence* 215). He therefore goes on to provide the following description:

Whoever has committed a murder, for example, feels guilty as soon as his passion is gone; he deserves, according to his own judgment, pain as requital [*Leid als Vergeltung*]; tormented by such thoughts, he wishes his deed could have been undone; he loathes his character. This state of mind is denoted by the expression “pangs of conscience” [*Gewissensbisse*]. (215–216)²⁸

It is important to see how these elements enrich the picture we were given in *Origin*. First, and especially crucial, is the idea that guilt as such involves the thought that one is *worthy of punishment*. Second, Rée adds here the idea that guilt involves a form of hatred of one’s own *character*.²⁹ As we saw, according to the earlier *Origin*, the experience of guilt as involving such loathing of character only appears in the quasi guilt feelings of the person who has recognized the unfreedom of the will, whereas here the loathing of one’s character is presented as part of the phenomenon of guilt itself. Third, the characterization of guilt as we have it here is silent about whether guilt requires belief in one’s free will; there is only the wish that the deed could have been undone, not the thought that one could have acted differently.

Let me now discuss Rée’s ideas about the origin and nature of punishment. In his discussion of the origin of the concept of requital or retribution (*Vergeltung*), which Rée equates with the

²⁸ Rée also adds that conscience can also speak to us in advance, before we perform the immoral action (it is, in my terminology, prospective), even though in the throes of passion one can barely hear it.

²⁹ Rée, however, does not develop the notion of character and its relation to the feeling of guilt in the text.

feeling of justice (*Gerechtigkeitsgefühl*), he repeats an analysis already articulated in *Origin*. According to this line of thought, the idea that the criminal *deserves* punishment for wrongful deeds is based on an error. Originally, Rée claims, the purpose of punishment was not to requite but to *deter*.³⁰ Punishment was adopted for the sake of its perceived beneficial consequences for the community. It was thus not backward-looking but forward-looking, implemented as a security measure (*Emergence* 318) and executed by “particular people, labeled as ‘the authorities’” (*Origin* 114), whose function required them not to reflect upon or proclaim in public the original purpose of punishment but merely to “establish that the accused person actually committed the crime he is charged with, of checking in the legal statutes what punishment is set down for the crime in question, and of carrying out this punishment” (114). Now, given that the entire punitive procedure did not explicitly display and thus inform or remind people of the deterring function of punishment, this original function was *forgotten*. As a result, “the judges, the accused, and the spectators” came to think that punishment is “retribution for the past”:

The judges merely assert that, under these circumstances, this punishment must be inflicted for theft, and they carry it out. It seems therefore that they are punishing not to prevent other thefts but because of the theft already committed; they seem to be taking retribution for the theft already committed. (*Origin* 114)

There thus arises an “optical illusion” (*Emergence* 319), and we acquire the belief that punishment has a retributive purpose, that the criminal *deserves* to be punished. Rée adds that the idea of retributive justice is also sown in us as children by means of education.

³⁰ It is possible that Rée was influenced here by Schopenhauer’s views on the matter: Schopenhauer, too, believed that the function and justification of punishment are *deterrence* and that it is thus *future*-oriented (see WWR I:347ff.).

Thus, though the punishments we receive as children are in actual fact meant to prevent the recurrence of our misbehavior, we are nevertheless told: "You are being punished because you have done this," so that "from childhood onward we have the impression that punishment is retribution, and this feeling is developed further by everything we see and hear about legal punishment" (*Origin* 115).

One should pause here to reflect on Rée's ingenious suggestion. The claim is that punishment has a visible as well as an invisible side (*Emergence* 319). The visible side concerns the forbidden action and the pain that is imposed on the culprit as punishment. The invisible side concerns the original purpose for the sake of which the punitive practice was adopted in the first place. This latter side, Rée claims, exists only in the minds of those who originally grasped the future-oriented goal of punishment but not in the minds of most men. But it is precisely the visible side that makes the most impression on people's minds (319) and thus, in the course of time, creates the illusion that punishment's role is retribution.

This account, however, seems only to explain why the original purpose of punishment is forgotten; it does not adequately explain how the *specific* (mistaken) notion emerged that punishment is retributive. The mere fact that A, a prohibited act of some kind, is seen to be followed by B, the punishment, can give rise to the notion that there is *some* connection between them, but it does not in itself seem to settle the question of *what* exactly the connection is.

But this criticism, I believe, fails to appreciate the radical nature of Rée's explanation of the notion of retributive punishment. For according to Rée, I wish to suggest, the belief that a crime or some other form of misbehavior deserves punishment is *nothing* over and above the thought that every crime must be followed by a punishment—a thought that we acquire as a result of mental conditioning practiced on us in education and socialization from very early on. The practices of visiting punishment upon the criminal or disciplining the child give rise in the course of time to a mental *habit* where people come to *associate* deeds regarded as wrong in

some sense with punishment and so start to *expect* crimes and misbehavior of different kinds to be followed by punishment. As a result of this process of mental conditioning, upon witnessing a crime, “one involuntarily has the feeling that some retribution must follow upon it—that is, the feeling of justice” (*Origin* 114). And if no punishment is forthcoming, one’s mental habit might motivate one to take action and punish the wrongdoer oneself or demand punishment (from the authorities or divinities) so as to scratch the mental itch that forms as a result of the unsatisfied expectation. As Rée puts it, “Formed in this way, the mind now *demand*s pain whenever an action takes place with which the consequence of pain is constantly thought together” (*Emergence* 320; emphasis added). We then say that the criminal *should* be punished.

We can see here, in his treatment of punishment, how Rée again employs his typical schema of explanation somewhat mockingly summarized by Nietzsche in his *Genealogy* (GM I:2): a practice is adopted because of its perceived beneficial consequences or *utility*, its original purpose is then *forgotten*, but its adoption generated a *habit* whereby one element of the practice becomes psychologically *associated* with another and seems to be inextricably tied to it. This tight association then generates a *new interpretation* of one of the two elements or of the relation that obtains between them. This is, of course, very reminiscent of Hume’s approach in philosophy, especially of his account of causation, where the idea of a necessary connection arises precisely out of a mental habituation that consists of experiencing a constant conjunction of similar events. Rée’s explanation of retributive justice could be seen as following Hume’s schema, where the constant conjunction involves crime and punishment and the result is the expectation that the one should follow the other, an expectation that is expressed with the view that the crime *deserves* to be punished.

In response to the challenge that in the idea of retribution we have *more* in mind than the mere thought that punishment must follow upon the crime, Rée can respond that it is not clear what

exactly this “more” is. The idea that the criminal deserves her punishment is clearly different from the idea of justice as deterrent or preventive, but what *positive* content this idea has it is hard to say. We say that the criminal *deserves* to be punished, or that *justice* requires that she be punished, and that if she is not punished, she *ought* to be punished regardless of considerations of deterrence or prevention. But Rée, presumably, would insist that these ideas do not get us any farther than the idea that the criminal deserves punishment—an idea that is merely expressive of how strongly we associate the crime and the punishment and so expect the one to be followed by the other and are even willing to take action to make that happen.

Don’t we, however, believe that a criminal deserves punishment only on the condition that she could have acted differently, that she was *free* at the time of action? It is important to emphasize that according to Rée’s analysis in *Emergence*, the idea that one deserves to be punished does not originally arise on the basis of the supposition that the culprit could have acted differently. The notion of desert—in other words—is not, or at least not originally, premised on the idea of leeway freedom. This marks a significant change in Rée’s account of punishment from the time of *Origin*.³¹ While the general account of punishment is the same in both works, in the earlier *Origin*, Rée also holds that the spirit of justice (*Gerechtigkeitsgefühl*), the thought that the criminal deserves punishment, arises out of *two* errors: the error of believing that the purpose of judgment is retributive *and* the error of free will (*Origin* 115).³² We believe, according to *Origin*, that the criminal deserves punishment because we also believe she could have acted differently and is therefore

³¹ In his discussion of Rée’s view on the bad conscience, Janaway treats the discussions in *Origin* and *Emergence* as virtually the same and thus fails to notice this important change in Rée’s view (Janaway 2007, 84–87).

³² In fact, Rée is not perfectly consistent on this point, for in an earlier remark, as we saw, he claims simply that “one involuntarily has the feeling that some retribution must follow upon it—that is, the feeling of justice” (*Origin* 114), as if this feeling of justice does not presuppose any belief in the agent’s free will at all.

responsible for her deeds. In contrast, in the later *Emergence*, no mention of free will is made in Rée's account of punishment or anywhere else in the book, including in his account of conscience. Why this change of mind? Is this a mere oversight on Rée's part, or is there a principled reason behind this revision?

I will suggest an answer to this question in the next section but wish now to return to Rée's account of the phenomenon of pangs of conscience in *Emergence*. With the account of punishment and our sense of justice (desert) now in place, Rée adds that just as we are educated to associate a certain moral value such as "reprehensible" or "categorically forbidden" with certain kinds of action (along the lines of explanation articulated in *Origin*), so are we at the same time habituated into associating the idea that these same actions are "punishable" or "deserve to be punished"; these associations all "grow out of the same stem" (*Emergence* 320). We can thus see how the central elements of the bad conscience form: when we ourselves act immorally, once the passion of the act subsides, we involuntarily judge that we have done a reprehensible deed, feel the negative feeling associated with the bad action, and consider ourselves deserving of punishment for it (330). We not only judge our action "bad" and feel bad for what we have done, as the account of *Origin* explains, but we also feel that we are worthy of punishment, that we *should* be punished. Importantly, at the time of *Emergence*, it thus seems, it is not a necessary condition of the feeling of guilt that we judge that we could have acted differently, that we possessed leeway freedom at the time of action. Nevertheless, and even though we don't assume free will in our experience of guilt, the fact remains that since freedom of will is an illusion, this implies for Rée that guilt is unjustified (I will return to this).

The feeling of justice—the demand that the crime be followed by punishment—also has the interesting consequence that it can serve to explain a familiar aspect of the feeling of guilt or bad conscience that Rée does not explicitly bring up, namely, that the guilty person might not only think he deserves punishment but also welcome

it and even actively *seek it out* if he realizes it is not forthcoming. Because of the strong psychological link between crime and punishment, the fact that one is not followed by the other can, quite literally, drive a guilty person out of his mind and impel him to submit himself before the punitive authorities. The punishment can then relieve him of the unbearable psychological tension; it provides a fulfillment of the intention of expectation, to put it in the terms of phenomenology.

Is, however, the thought that one deserves punishment an essential element of the feeling of guilt? Was Rée right to broaden the account of guilt given in *Origin* in this way? It definitely does not seem to be true that all instances of guilt involve the thought that one deserves to be punished (a fortiori, not all cases of guilt give rise to an active seeking out of punishment when it does not arrive). For example, I might feel guilty for pressing my son too hard to work on his math homework, causing him to break down and cry, without feeling the need to be punished for my parental strictness. It seems to me that the thought that one deserves to be punished arises, though probably not invariably, in cases of guilt for misdeeds for which there is an extant, legally institutionalized, and *severe* punitive practice, that is, in cases of *severe crimes*. If this is correct, an explanation might be that it is in the case of such misdeeds that the described psychological association could be expected to become entrenched in the most impressive manner, given the severity of the customary punishment.

Let me now turn to another respect in which the account of remorse in *Emergence* goes beyond the one in *Origin*, namely, that it provides us with a helpful set of distinctions that further sharpens our understanding of the phenomenon of guilt (*Emergence* 330–331). A first distinction Rée makes, which plays no significant role in his analysis, is that between pangs of conscience (*Gewissensbisse*) and consciousness of guilt (*Schuldbewusstsein*), where the difference between the two is merely that the latter signifies longer-lasting pangs of conscience (330; cf. Schopenhauer's view in WWR

I:364–365). More important distinctions are introduced under the category of regret (*Bedauern*).

Under “regret,” Rée distinguishes three different phenomena he claims are often confused. First, there is what he calls “ethical regret” (*sittliches Bedauern*; *Emergence* 330). This is the phenomenon of the pangs of conscience we have just been analyzing—the genuine phenomenon of guilt. Second, there is what Rée calls “unegoistic regret” (*unegoistisches Bedauern*), with which ethical regret is most frequently confused (331). This basically refers to how Darwin understands the phenomenon of bad conscience according to Rée’s reading, namely, as the feeling of the unsatisfied instinct of compassion which appears on one’s mental scene and makes its protestations after the immoral, egoistic action has already been performed and the passion that motivated it has as a result subsided. Thus, after having committed murder, the murderer might be struck not with ethical regret but with “compassion complaining about what vengefulness or avarice has performed” (331). As we saw, Rée is of the view that this kind of feeling does *not* amount to pangs of conscience, because it does not differ in its essence from the feeling we experience whenever one of our instincts remains unfulfilled and protests its lack of fulfillment. Thus, Rée goes on to explain, it is possible to feel unegoistic regret even in a society where we have been conditioned to feel ethical regret about acts of compassion (for the latter, Rée gives the example of “the cannibal”; 330), a society where lack of compassion counts as something *praiseworthy* (332).³³ Alternatively, it is possible to feel unegoistic regret when we have not been conditioned to judge compassion (or its lack) in *any* way, as either praiseworthy or reprehensible (332). Finally, and this is the case in the “higher cultures,” it is possible to feel both ethical regret and unegoistic regret at the

³³ This view of Rée’s is indicative of the greater relativism he espouses in *Emergence*, one that goes beyond the views articulated in *Origin*. This is another aspect in which his thought underwent change.

same time—to feel “double regret” (332). With this, we have Rée addressing a question I raised earlier regarding the relation between the unsatisfied drive of compassion and remorse: Rée’s (more mature) view in *Emergence* is that they could be experienced simultaneously.

The third kind of regret Rée distinguishes is what he calls “egoistic regret” (*egoistisches Bedauern*), where one is “seized by fear of punishment” (330; cf. Schopenhauer’s “spurious conscience”). This distinction helps us see that the tormenting thought that one deserves punishment, which makes up *part* of ethical regret, is different from the fear of punishment at issue here: one can fear punishment without believing that one deserves it because one does not believe one has done something wrong. But the converse is possible as well: one can believe that one deserves punishment without fearing the punishment itself. Finally, it is also possible to fear the punishment *and* believe one deserves it. Thus, an interesting result of this analysis is that the phenomenon of bad conscience or ethical regret can present itself—and perhaps usually presents itself—in impure form, that is, as an admixture of two or three kinds of regret. Thus, one can feel bad about what one has done, think that one deserves punishment for it, fear the punishment, *and* feel the pain of unsatisfied compassion.

IV. The Illusion of Free Will

As we saw, in the earlier *Origin*, Rée expresses the view that guilt presupposes the thought that one could have acted differently, the idea that generates the illusion of the freedom of the will. In contrast, in *Emergence*, in his description of punishment as retribution as well as in his discussion of ethical regret, this component is missing. Rée seems to have changed his mind, and free will is not presented anymore as a necessary element that figures into the thought that punishment is deserved or into the experience of guilt.

In his short and last work from the same year as *Emergence, The Illusion of Free Will: Its Causes and Consequences* (1885), the concept of free will reappears and is discussed on the background of a new explanation Rée provides for our belief that we are free. As I will show, an explanation of Rée's change of mind can be found in this latter work.

As we saw, in the earlier *Origin*, Rée argues that the reason we believe we are free is that we think a certain capacity that we possess could have also been exercised at the time of the prohibited action. Here, in *Illusion*, this is replaced by a different hypothesis, according to which the explanation for our false belief in free will involves our inability to perceive the minute and complex causal relations that govern what we think and feel and how these give rise to decision and action. He explains in the first section of the work that "we do not perceive the causes through which our will is determined, and consequently think that it is not conditioned causally by anything at all" (*Illusion* 364). A bit later in the text, Rée observes that it is this lack of awareness of the sufficient causes of our actions that makes us think that "we could have just as well willed differently and acted differently" (367). It thus might seem that here, too, just as in *Origin*, Rée's view is that the (false) belief in leeway freedom serves as the condition for the thought that one is responsible and therefore guilty (when one performs a blameworthy deed).

This, however, is not the case, for in the second section of the work, he claims that "we hold ourselves and others responsible without thinking about the problem of the freedom of the will" (369). And he later adds that as a result of the processes of acculturation (considered earlier), we come to judge a wrong action as reprehensible involuntarily and do not raise the question of "whether the action is determined causally or not," and this regardless of whether we are the agents of the action or are merely its observers (372)—this just does not come to our mind. As Rée claims, it is mostly *philosophers* who ask whether actions are causally determined or not, not the rest of humanity (372). This is one

reason, I suggest, Rée drops freedom of the will as a component of guilt in *Emergence*: we feel guilty without inquiring into or even knowing anything about determinism and the problem of free will. The experience of guilt does not subjectively presuppose the belief in free will.

Another possible reason for Rée's change of mind, which perhaps accounts for the first reason, is the following. As we saw, in *Emergence*, guilt is conceived as involving the thought that one deserves to be punished. But, as we also saw here, this thought, that the criminal deserves punishment, is accounted for psychologically without making use of the notion of free will. When the "original" authorities founded the practice of punishing a criminal, they were interested, Rée holds, not in retribution but in deterrence. They therefore did not need to take, and did not take, any interest in the metaphysical question of whether a criminal's will is determined or not but merely aimed to disincentivize future occurrences of criminal behavior on the part of the criminal himself and on the part of others. The philosophical considerations regarding the freedom of the will thus do not enter as constitutive elements into the formation of the psychological habit of associating crime with punishment. They thus do not take root in one's consciousness and do not enter one's mind in the experience of guilt. Similar conditions obtain in the process of education: when a parent punishes a child for what he or she has done, questions about the child's free will do not arise and are never presented to the child as a presupposition that must hold for the punishment to be administered. Consequently, the psychological knot that takes shape in the child's mind passes through *two* nodes, the crime and the punishment; it does not pass through a third.

This strongly suggests that for Rée at the time of *Emergence* and *Illusion*, the feeling of justice in its origin has *nothing to do* with the freedom of the will, that is, freedom from causal determinism. If this is correct, then we can see how Nietzsche, again, is off the mark

in his criticisms of Rée in the *Genealogy*. In the second essay of the *Genealogy*, Nietzsche writes:

Have these genealogists of morals had even the remotest suspicion that . . . punishment, as requital, evolved quite independently of any presupposition concerning freedom or non-freedom of the will? . . . The idea, now so obvious, apparently so natural, even unavoidable, that had to serve as the explanation of how the sense of justice [*Gerechtigkeitsgefühl*] ever appeared on earth—"the criminal deserves punishment *because* he could have acted differently"—is in fact an extremely late and subtle form of human judgment and inference. (GM II:4)

As Janaway argues (2007, 86–87), there is very good reason to think that Nietzsche has Rée in mind in this passage, but Janaway fails to notice that Nietzsche's criticism could be regarded as valid only with respect to Rée's *Origin* but not with respect to his later writings, which Nietzsche read even *before* the publication of his *Genealogy* in 1887. In actual fact, Rée's more mature view is precisely that punishment as requital does *not* involve in its origin any ideas about the "freedom or non-freedom of the will."

Rée's mature view, however, raises at least a couple of related issues. The first is this. On the one hand, as we saw, he holds that the problem of the freedom of the will just does not arise when we regard others or ourselves as guilty of misdeeds. Nevertheless, he adds, importantly, that if we discover—say, with the help of some philosopher—the necessity of all action, we will not consider anyone as *deserving* punishment for his or her misdeeds or praise and credit for his or her good deeds (*Illusion* 370, 372, 373). Moreover, we will not consider *ourselves* as guilty for our misdeeds or feel guilty (373–374), the latter provided that our judgment regarding the unfreedom of the will can overcome the habits of thought we acquired as children, whereby we became accustomed to feel guilty for our misdeeds independently of any metaphysical assumptions.

But why should that be the case? *Why* would we, or should we, cease to think that anybody deserves blame or punishment or merits praise once we acquire metaphysical insight into the (alleged) truth of determinism? If our practices and psychological habits are *not* grounded in a belief about our free will, then why would we or should we reconsider these practices when we grasp the falsity of this belief? Why would a structure collapse if the beam that is pulled out from under it does not serve to support it in the first place? It seems our practices of praise and blame could only suffer this collapse if they *are* founded, if only implicitly, on the belief that we *do* have free will.

It is possible, though Rée does not say this, that only very recently, for us moderns, as a result of certain historical processes, the ideas of deserving punishment on the one hand and free will on the other have become intertwined in such a way that we think the one grounds the other, so that when we realize that free will is an illusion, we also reject, or come to think we should reject, the notion of desert. Another possibility is that perhaps the philosophical interrogation and rejection of free will make one grasp certain logical connections between freedom and responsibility not visible before, so that one only retroactively, subsequently to the rejection of free will, comes to see that responsibility actually presupposes free will—it presupposed it all along—and so comes, with the rejection of free will, to reject responsibility as well, or to believe it should be rejected.

The second issue is this. On the one hand, as we saw, Rée says that because we don't clearly perceive in its entirety the causal nexus in which we are caught, we non-philosophers think our actions are not causally determined, that our will "is not conditioned causally by anything at all" (*Illusion* 364), and therefore think, "I could have acted differently." On the other hand, Rée claims that when we hold ourselves or others responsible or judge others or ourselves guilty, we *don't* concern ourselves with the question of whether the will is determined or not. How can we reconcile these ideas?

There is, of course, no direct contradiction here, for it is possible to hold that because we cannot see into the nitty-gritty of the causal network, we indeed believe that people could have acted differently, but that this belief plays no justificatory role (or any other role) in our holding people responsible for their actions, and, indeed, Rée does not seem to ascribe any such role to this belief in *Illusion*. This, however, is rather implausible, in my view, because it does not clarify what the role of the “I could have acted differently” thought *is*, if its function is *not* to serve as such justification. One possible way to think about this is to sever the connection between the freedom of the will as a technical, philosophical problem, on the one hand, and the “could have acted differently” thought, on the other. Once the connection is severed, we can hold that the “could have acted differently” thought *does* justify people’s holding themselves and others responsible and so grounds guilt and punishment but that this “could have acted differently” is not grasped clearly by most people, that is, by non-philosophers, and therefore does not involve seeing this thought’s connection to the problem of free will. As Rée claims, most people do not see the problem in its “general form (is the will free from causal laws or not?)” (*Illusion* 367). If we opt for this solution, then it turns out Nietzsche perhaps was *partly* right in this criticism of Rée in the *Genealogy*: Rée presumably ties the idea of desert to the “could have acted differently,” but—and here Nietzsche goes wrong—this idea is not equivalent for him with the philosophical problem regarding the freedom or unfreedom of the will. Nietzsche, in his criticism, conflates the two.

I believe Rée, on this reconstruction, is trying to say something that is fundamentally true, namely, that most people (non-philosophers or even philosophers in their non-philosophical moments) do not consider the metaphysical questions concerning freedom of the will when they assign praise or blame or hold people responsible. So when we feel guilty, we might think we deserve punishment because—we think—we could have acted differently, but this latter thought is not one reached as the conclusion of some sort

of metaphysical investigation regarding the freedom of the will or the nature of causality. The view that these two ideas are connected and that the “could have acted differently” can only be justified if the will is free from all causality is a substantial philosophical thesis that is neither common among the non-philosophical nor uncontroversial among professional philosophers.

A puzzle remains: why *would* the “could have acted differently” thought be part of our experience of guilt or part of our justification of punishment? As we saw, according to Rée, the origins of punishment and the process of education do not depend on any such idea. Is there reason to think they involve the notion of being able to act differently in such a way that it gets impressed upon one’s mind and thus determines one’s psychological associations? It doesn’t seem so. Where is its origin to be found, then?

In chapter 5, I shall examine Nietzsche’s ideas about guilt and its genealogy and wonder in what way and to what extent Nietzsche’s thoughts go beyond and surpass Rée’s. Should we prefer Nietzsche’s account to Rée’s? Should we, rather, see the two as complementary?

5

Nietzsche

The Genealogy of Guilt

Introduction

What does Friedrich Nietzsche have to add to Rée's naturalistic understanding of guilt? This question might sound deliberately provocative or even, to some, heretical. In contrast to Rée's analysis, Nietzsche's genealogy of guilt in the second essay of *On the Genealogy of Morals*—entitled 'Guilt,' 'Bad Conscience,' and the Like"—is labyrinthine in nature and proceeds in meandering lines of thought that go back and forth and in different directions, to the dismay and (occasional) despair of the reader. So why should we wrestle our way through the disorienting thicket that is the *Genealogy* rather than be satisfied with Rée-istic clarity? Does Nietzsche expose gaps, inconsistencies, or falsities in Rée's account? I have already argued in chapter 4 that two of his central criticisms of Rée miss their mark.

Granted, Rée leaves some issues unresolved. But does Nietzsche provide answers to these questions, solutions to these problems?

In this chapter, I will reconstruct Nietzsche's genealogy of guilt in detail. One of the main claims I will be making is that Nietzsche's genealogy of guilt can be seen to go deeper than Rée's in that it provides us with a genealogy of social and mental structures that Rée's genealogy presupposes. On the other hand, as I will argue, at various crucial junctions, Nietzsche can be read as helping himself to a Rée-ian form of explanation. Before I turn to Nietzsche's genealogy of guilt in his *Genealogy* and other writings, I examine his

earlier critique of Schopenhauer in *Human, All Too Human* and his critique of the idea of *causa sui* in *Beyond Good and Evil*.

I. Nietzsche's Critique of Schopenhauer's Argument

It is well known that Schopenhauer exerted a great influence over Nietzsche's thought and that Nietzsche critically engaged with his ideas throughout his career, starting especially with his 1878 *Human, All Too Human*, where he asserts his intellectual independence and subjects to relentless criticism many ideas of his erstwhile philosophical mentor. It is at that time that Nietzsche began his occupation with moral psychology and the critique of morality,¹ including his persistent attacks on the ideas of free will and moral responsibility—attacks that took various forms and lasted till his very last writings. Thus, one argument Nietzsche makes is that given the strict determinism that obtains in nature, the “curvings, twistings and breakings” of a waterfall are all “necessary, every motion mathematically calculable.” And Nietzsche continues: “So it is too in the case of human actions. . . . [But the] actor himself, to be sure, is fixed in the illusion of free will” (HH 106). From this necessity of all events, it follows, Nietzsche holds, that no person could have acted differently from how he actually did (HH 105). He then concludes, “He who has fully grasped the theory of total irresponsibility [*Unverantwortlichkeit*] can no longer accommodate the so-called justice that punishes and rewards. . . . For he who is punished does not deserve the punishment . . . for he could not have acted otherwise” (105; translation slightly altered). Briefly put, free will in the sense of being able to act differently, or leeway freedom, is false, and consequently, no person is morally responsible. Here, as in

¹ Even though Nietzsche himself claims that his “campaign against morality” started in the later *Daybreak*. See EH, *Books, Dawn*, 1.

other places in HH, Nietzsche presupposes determinism and does not argue for it, in part presumably because he takes it to have been established by Schopenhauer and his predecessors.

But while many scholars have drawn attention to Nietzsche's criticism of free will and moral responsibility in his middle period (see Abbey 2000; Franco 2011), in a section of great interests, HH 39, Nietzsche is putting forward a different argument which specifically attacks Schopenhauer's transcendental argument from guilt to freedom. Nietzsche begins the section by providing the reader with "the principal stages in the history of the sensations by virtue of which we make anyone responsible for his actions" (HH 39; translation slightly altered). Thus, Nietzsche explains, in a manner that to some extent follows Rée's genealogy, that at first we call the consequence of actions "good" (*gut*) or "evil" (*böse*), then forget the origin of these terms and start to think that the actions themselves are inherently good or evil, proceed to consider the actions' motives as good or evil and the actions themselves as morally ambiguous, and conclude by considering "the whole being of man out of whom the motives grow" as good or evil. The final stage in this progression, Nietzsche claims, is that "one finally discovers that this being [*Wesen*], too, cannot be responsible [*Verantwortlich*], inasmuch as it is altogether a necessary consequence and assembled from the elements and influence of things past and present" (HH 39; translation slightly altered). It thus follows, Nietzsche maintains, that responsibility is an error "which rests on the error of freedom of will." Here again, we have the familiar argument that moves from complete determinism and the denial of free will to the denial of responsibility. But Nietzsche immediately proceeds to add (and I quote him at some length):

Schopenhauer concluded otherwise, thus: because certain actions bring after them a feeling of *displeasure* [*Unmuth*] ("consciousness of guilt" [*Schuldbewusstsein*]), there must exist responsibility [*Verantwortlichkeit*]; for there would be *no ground* for this feeling

of displeasure if not only were all the actions of man determined by necessity—which is in fact the case, a view also held by this philosopher—but man himself acquired his entire *being* [*Wesen*] with this same necessity—which Schopenhauer denies. From the fact of that feeling of displeasure Schopenhauer believes he can demonstrate a freedom which man must have had somehow, not in respect of his actions but in respect to his being: freedom to *be* thus or thus, that is to say, not to *act* thus or thus. From the *esse*, the sphere of freedom and responsibility, there follows in his opinion the *operari*—the sphere of strict causality, necessity and lack of responsibility. That feeling of displeasure appears to relate to the *operari*, to be sure—to that extent it is in error—in truth, however it relates to the *esse*, which is the deed [*That*] of the free will, the primary cause of an individual's existence: man becomes that which he *wills* to become, his willing precedes his existence. Here the erroneous conclusion is drawn that from the fact of a feeling of displeasure there can be inferred the justification, the rational *admissibility* of this feeling of displeasure; and from this erroneous conclusion Schopenhauer arrives at his fantastic concept of so-called intelligible freedom. But a feeling of displeasure after a deed is absolutely not obliged to be rational; on the contrary, it cannot be, since it rests precisely on the erroneous presupposition that the deed need *not* have taken place of necessity. Thus: it is because man *regards* himself as free, not because he is free, that he feels remorse and pangs of conscience [*Reue und Gewissensbisse*]. (HH 39)

We can readily see how Nietzsche captures the essentials of Schopenhauer's position in this passage, and we can thus appreciate how, in his engagement with Schopenhauer's view, Nietzsche is in effect confronting the entire tradition out of which Schopenhauer's view grows. In addition, we can notice how Nietzsche's criticism differs from Rée's. As we saw in chapter 4, Rée dismisses Schopenhauer's metaphysical view in preference to Darwin's

“simpler” view (Rée, *Origin* 92). But in this section, Nietzsche engages Schopenhauer in a more direct manner. What exactly are Nietzsche’s objections to Schopenhauer’s view, however?

Nietzsche’s attack on Schopenhauer’s argument in HH 39 involves two related yet distinct lines of criticism. The first, which Nietzsche raises at the very end of the quotation and which basically reiterates Rée’s view in *Origin*, is that people feel guilty because they erroneously think they are free. The erroneous assumption of freedom, Nietzsche holds, grounds the feeling of guilt, which then gets rationalized transcendently by philosophers such as Schopenhauer, thus giving rise to what Nietzsche thinks are “fantastic” metaphysical conclusions. Once this assumption of freedom is exposed for its error, the feeling of guilt should be seen as lacking any ground and could in principle be “disaccustomed to,” as Nietzsche goes on to claim without great elaboration, besides noting that such feelings are a “very changeable thing, tied to the evolution of morality and culture” (HH 39).²

This attack on Schopenhauer’s argument, however, merely presents us with the *modus tollens* of Schopenhauer’s argument: while the latter, according to Nietzsche, argues from the justifiability of the feeling of guilt to freedom, Nietzsche argues from the denial of freedom to the denial of the justifiability of the feeling of guilt. This threatens to put the two thinkers at a dialectical standoff. When this is looked at more closely, however, a deeper issue arises: Nietzsche’s attack relies, it seems, on the denial of free will in the realm of appearances—the freedom he criticizes, for example, in HH 106. But Schopenhauer would be in complete agreement with Nietzsche on this score. Thus, Nietzsche’s denial of free will

² See also HH 107 and HH 133. As we shall see, Nietzsche in his genealogy attempts to show how the feeling of guilt comes about and develops and therefore cannot be seen as belonging to humans as such or to humans’ metaphysical nature. It is worth noting that here Nietzsche seems to argue that guilt should be rejected because it is unjustified—that is, because of its faulty epistemic status. Later in his career, Nietzsche came to hold the view that the truth or falsity of a judgment is not a necessary reason for rejecting it but whether it promotes or does not promote “life” (see BGE 4).

on the basis of which he wishes to deny the justifiability of guilt fails to engage Schopenhauer at the proper level, for it does not aim at Schopenhauer's conception of *intelligible* freedom. Schopenhauer can thus agree that our actions are determined but hold that this shows nothing about our noumenal, intelligible freedom, and thus nothing about the justifiability of our feelings of guilt. As we saw, Nietzsche, of course, asserts in his short history of morality that "one finally discovers that this being, too, cannot be responsible, inasmuch as it is altogether a necessary consequence and assembled from the elements and influence of things past and present" (HH 39), but this merely amounts to an unargued-for rejection of Schopenhauer's claims about the intelligible character—the being of any individual person—as lying outside of space and time and thus outside the scope of "influence of things past and present." It is true that Nietzsche addresses the notion of the metaphysical world directly in HH 9, but he explicitly admits in this section that the "absolute possibility of it is hardly to be disputed," even after we discover all reasons for belief in such a world were based on "the worst of all methods of acquiring knowledge," namely, on "passion, error and self-deception" (HH 9). But this only forces us to raise the question again: what exactly is Nietzsche's criticism of Schopenhauer?

The second criticism one can extract from Nietzsche's comments cuts deeper. According to this line of attack, the problem with Schopenhauer's argument lies in its premise according to which our feeling of guilt must be rational or justified—a premise that Schopenhauer does not argue for. While Schopenhauer, as we saw, does not present his argument in this specific way, he does assume that our empirical guilt and what it announces (our responsibility) are justified, and Nietzsche's reconstruction of the argument exposes this crucial premise, without which Schopenhauer's argument cannot go through, a premise that, in a way, underlies the entirety of the metaphysical tradition, for in all three cases examined, guilt—ontological or empirical—is assumed to be justified or imputable in some sense. Our guilt, even if it *feels* justified,

is not self-evidently justified—Nietzsche claims—and to erect an entire metaphysical construction on the basis of this illicit premise is nothing but a strong indication of how philosophy has been bewitched, since the days of Plato, by morality. To paraphrase Ludwig Wittgenstein, from Nietzsche's perspective, an entire cloud of philosophy is condensed here into a drop of psychological mistake: I feel or experience something, *so* it must be justified, all the more so when the feeling is a moral one, the feeling of guilt. In contrast, Nietzsche thinks that "the sting of conscience is, like a scorpion stinging a stone, a piece of stupidity" (WS 38), which as such *lacks* justification.³ Schopenhauer's—and, in related ways, Schelling's and Kant's—entire argument collapses, and we are left with no reason to infer the existence of anything like intelligible freedom.

Why, however, does Nietzsche *reject* Schopenhauer's implicit assumption that our feeling of guilt is justified? Lack of justification on Schopenhauer's part should lead Nietzsche to a suspension of judgment rather than rejection. It is clear that Nietzsche goes beyond suspension and rejects the assumption, for right after claiming that the feeling need not be rational, Nietzsche adds, "on the contrary, it cannot be, since it rests precisely on the erroneous

³ In the original German, the metaphor concerns not a scorpion but the bite of a dog. In Hollingdale's translation, in order to align itself with the English expression "sting of conscience," the biting becomes a stinging, but, alas, the dog becomes a snake, which is not known to sting but to bite. I therefore transformed the snake into a scorpion. The German *Stein* is kept as "stone" both in Hollingdale's and in my translation. The question remains, however, of how to unpack the metaphor at work here. It is, of course, silly of the scorpion to sting the stone, given that the stone poses no threat. But surely Nietzsche is not trying to say that it is unreasonable of us to lacerate ourselves with feelings of guilt because we pose no threat to ourselves. Moreover, the scorpion attacks an entity other than itself, while we aim the sting of conscience at ourselves. Nietzsche's point seems to be that in both cases, we are dealing with a natural phenomenon that is based on a cognitive error: the scorpion mistakes the stone for an enemy, and we mistakenly believe our guilt is justified since we are free and could have acted differently; in both cases, a violent reaction is based on an error. In addition, in both cases, a naturalistic explanation for this error could in principle be given; the scorpion's cognitive apparatus is falsely triggered by an entity that sufficiently resembles animals against which it is constituted to defend itself, and our feelings of guilt arise out of a complex psychological mechanism that in addition presupposes the belief in our free will.

presupposition that the deed need *not* have taken place of necessity” (HH 39). But this is just to reiterate the denial of the phenomenal freedom of the will.

An additional reason to reject the assumption can be found in the following. As Nietzsche claims in *The Wanderer and His Shadow* 7, apropos Epicurus, there could be *various explanations* for our feeling of guilt which do not amount to its justification. As he puts it:

A *multiplicity* of hypotheses, for example as to the origin [*Herkunft*] of the sting of conscience [*Gewissensbisse*], suffices still in our own time to lift from the soul that shadow that so easily arises from a laborious pondering over a single hypothesis which, being the only one visible, is a hundredfold overrated. (WS 7)

Nietzsche does not explicitly state what this “single hypothesis” is, but it is not far-fetched to assume that what he has in mind here is the idea, criticized in the *Genealogy* but already broached here, that the conscience is “the voice of God in the heart of man” (WS 52). Nevertheless, Nietzsche’s objection applies equally well to Schopenhauer’s justification of guilt: there could be many non-justifying explanations for the feeling of guilt. Moreover, at the time of HH, Rée had already provided one such detailed explanation.

Why should, however, the plurality of possible alternative hypotheses lead us to reject Schopenhauer’s? A plurality of (equally) viable explanations should, again, only lead us to be skeptical of each, not to reject any. Rée’s consideration of simplicity, I think, cannot be easily invoked here, for a naturalistic account (such as Rée’s, or Nietzsche’s—as we shall see) is not clearly simpler than the Schopenhauerian one. So far, then, we have only been given reason to be skeptical of Schopenhauer’s argument—to put it on *ice*, to use Nietzsche’s expression (EH, Books, Human, 1). It is Nietzsche’s criticism of the notion of *causa sui* that delivers, I think, a powerful

blow against Schopenhauer's account as well as against the metaphysical tradition as a whole.

II. Nietzsche's Critique of the *Causa Sui* Argument

Later in his career, in *Beyond Good and Evil*, Nietzsche criticizes free will and responsibility on the basis of his rejection of the idea of *causa sui* which he takes to be "fundamentally absurd" (BGE 15).⁴ As he puts it in BGE 21:

The *causa sui* is the best self-contradiction that has been conceived so far, it is a sort of rape and perversion of logic; but the extravagant pride of man has managed to entangle itself profoundly and frightfully with just this nonsense. The desire for "freedom of the will" in the superlative metaphysical sense, which still holds sway, unfortunately, in the minds of the half-educated; the desire to bear the entire and ultimate responsibility for one's actions oneself, and to absolve God, the world, ancestors, chance, and society involves nothing less than to be precisely this *causa sui* and, with more than Münchhausen's audacity, to pull oneself up into existence by the hair, out of the swamps of nothingness.⁵

The thought at work here is that for moral responsibility to be possible, it must be possible for one "to bear the entire and ultimate responsibility for one's actions"; that is, one has to be the cause of oneself, since were one the result of other factors such as "God, the world, ancestors, chance, and society," one would not be such an

⁴ Nietzsche critically refers to the idea of *causa sui* already in his *Philosophy in the Tragic Age of the Greeks* in the context of his discussion of Anaxagoras. He there refers to it as a "contradictory concept" and as "forbidden" (PTAG 100).

⁵ Nietzsche goes on in the section to deny also the idea of the "unfree will," since it "amounts to a misuse of cause and effect" (BGE 21). I will not go into this here.

ultimate author of one's actions and could not then be held morally responsible for them. But since the idea of *causa sui* is "absurd," no one is truly morally responsible for their actions.⁶

One possible problem with Nietzsche's argument is that, as Christopher Janaway claims, it is missing a premise which states that "there can be free will only if there is a *causa sui*" (Janaway 2007, 115). Since, however, Nietzsche is arguing within a philosophical tradition where the idea of *causa sui* was taken to be necessary for free will and (consequently) moral responsibility and guilt, he does not and should not be expected to supply or argue for this additional premise. Indeed, what is not highlighted by commentators on the argument in BGE 21⁷ is that here, as in HH 39, Nietzsche is engaging an entire metaphysical tradition that, in order to make room for guilt and moral responsibility, has relied on the concept of *causa sui*.

In his later *Twilight of the Idols*, Nietzsche again criticizes the *causa sui* argument without mentioning the concept explicitly. He writes:

What is the only teaching *we* can have?—That no one *gives* people their qualities, not God or society, parents or ancestors, not even *people themselves* (—this final bit of nonsense was circulated by Kant—and maybe even by Plato—under the rubric of "intelligible freedom"). *Nobody* is responsible for people existing in the first place, or for the state or circumstances or environment they are in. . . . [N]obody is held responsible any more . . . *this* is how we begin to redeem the world. (TI, Errors, 8)

⁶ I emphasize that what is of importance is the notion of *moral* responsibility, for arguably, one can be a responsible person without being *causa sui* and without being responsible for one's actions in the stronger sense of moral responsibility. In fact, Nietzsche thinks that one can be nonmorally responsible in this sense, as he explains in the *Genealogy*. I will say more on this.

⁷ Thus, Walter Kaufmann, in his translator's note to BGE 21, reads it as targeting conceptions of freedom such as to be found in Jean-Paul Sartre's existentialism and thus ignores the historical background of Nietzsche's argument. But see also Leiter 2002, 88–91.

It is clear that Nietzsche is criticizing here Schopenhauer's *causa sui* notion of people giving their character to themselves in an act of intelligible freedom, a view that Schopenhauer himself, as we saw, thought to have found expressed already by Kant and Plato (which is why Nietzsche is referencing the latter two rather than Schopenhauer directly). As Nietzsche clearly sees, this "bit of nonsense" was used to make responsibility possible.

Why, however, is the notion of *causa sui* absurd, a "rape of logic"? Brian Leiter has rightly emphasized the importance of this argument in Nietzsche (Leiter 2002, 88–91), and Galen Strawson, influenced by Nietzsche's argument, provides a more conspicuous reconstruction of the absurdity:

- 1) You do what you do because of the way you are
So
- 2) To be truly morally responsible for what you do you must be truly responsible for the way you are—at least in certain crucial respects.
But
- 3) You cannot be truly responsible for the way you are, so you cannot be truly responsible for what you do.
And why can't you be truly responsible for the way you are?
Because
- 4) To be truly responsible for the way you are, you must have intentionally brought it about that you are the way you are, and this is impossible. (Strawson 2003, 229–230)

And Strawson continues to explain that this is impossible because to bring it about intentionally that you are the way you are, you must have had a "certain nature N in the light of which you intentionally brought it about that you are as you now are" (Strawson 2003, 230), but then the same line of questioning will arise with respect to this nature N, for which the person would have to be truly responsible in order to be truly responsible for his subsequent nature and the

actions resulting from it, and an infinite regress would ensue, thus showing that being responsible for the way you are and for what you do is, as we saw Nietzsche puts it, a “rape and perversion of logic.”

In fact, as we saw when considering John Atwell’s objection to the attribution of the idea to Schopenhauer (in chapter 3), the “perversion of logic” involved here could just as well be conceived as involving not an finite regress but the no less absurd idea of preexisting one’s own existence. For, we can say, in order to be responsible for who you are, you must have “pulled yourself up into existence,” but that means that *you* (not someone else!) must have already been there to do the pulling and so were already in existence before you came into existence, which is incoherent.

Arguably, however, the problem with the notion of *causa sui* has to do with time. We think that if something comes into existence at a certain time, then it did not exist before; thus, *first*, it does not exist, and *then* it does, which presupposes a temporal progression of events. But, as we saw, in the metaphysical tradition, the noumenal *That* or deed of self-creation is outside of time, and so it might be argued that the *causa sui* notion is not as absurd as it might otherwise appear. To this, I think, Nietzsche could respond that just to say (or think) that the deed of self-creation lies outside of time is not as a matter of fact to make it any more intelligible; it might seem as if we are saying (or thinking) something sensible here, but we are actually not, for the very notion of a deed implies for us (linear) temporality.

The Schopenhauerian argument thus is to be rejected: it relies on an unjustified premise (that the feeling of guilt is justified), there are alternative explanations for this feeling which do not involve any notion of intelligible freedom (explanations delivered by Rée and by Nietzsche himself), and, finally, it ultimately relies on the untenable notion of *causa sui*. Since we are not *causa sui*, guilt is unjustified.

III. Nietzsche's Genealogy of Bad Conscience and Guilt: The Basic Psychological Framework

I now turn to examine Nietzsche's genealogy of bad conscience and guilt. I focus on the second essay of the *Genealogy* but draw extensively on other writings. It is with this genealogy that Nietzsche provides his own hypothesis "as to the origin [*Herkunft*] of the sting of conscience [*Gewissensbisse*]" (WS 7) on the basis of which he aims to show that the feeling of guilt is "a very changeable thing, tied to the evolution of morality and culture and perhaps present in only a relatively brief span of world-history" (HH 39).

Now, the second essay of the *Genealogy* is arguably the most complex and challenging essay of the work, which is why disagreements among its interpreters are legion. The interpretative challenge is to piece together the various elements that are presented to the reader in nonlinear fashion in a manner that amounts to a progressive and intelligible genealogy, one that is philosophically plausible and illuminating. Instead of following Nietzsche's own order of presentation, I proceed to reconstruct Nietzsche's genealogy of bad conscience and guilt in five stages: (1) the basic psychological framework, (2) the prehistoric stage and the morality of custom, (3) the sovereign individual, (4) beyond the sovereign individual, and (5) Christian guilt.⁸ My interpretation of the second essay is based on the assumption—and here I depart to some extent from some readings—that Nietzsche does not distinguish bad conscience⁹ (*schelchtes Gewissen*, as in the title of the essay),

⁸ To some extent, my views on guilt have changed from the account I provide in Elgat 2017, chap. 5.

⁹ E.g., Leiter 2002; Reginster 2018. Leiter wishes to maintain a distinction between bad conscience and guilt. He claims that although Nietzsche seems to equate the two, they cannot be identical "for a variety of reasons: 1) the Greeks had bad conscience but did not suffer from guilt . . . ; 2) Nietzsche suggests [GM II:23–24] "bad conscience" could be enlisted in the service of normative ideals that would not produce guilt; and 3) the internalization of cruelty is not by itself sufficient to explain guilt" (Leiter 2002, 229n9). In response to (1): the case of the Greeks, which I will discuss, actually supports the identification of bad conscience and guilt, for here (GM II:23) Nietzsche says that the

guilt (*Schuld*, as in the title of the essay), guilt feelings (*das Gefühl der Schuld*, GM II:14), or consciousness of guilt (*das Bewusstsein der Schuld*, GM II:4) and for the most part treats these as equivalent.¹⁰ This is not to deny, however, that bad conscience and guilt undergo various transformations in the course of the genealogy that Nietzsche traces. Let me now start with Nietzsche's basic ideas about bad conscience and guilt.

On the background of Rée's genealogy, perhaps the first thing to notice is that in contrast to Rée, Nietzsche does not analyze human psychology in terms of a fundamental division between the egoistic and the non-egoistic drives¹¹ and consequently does not posit the existence of any such drives at the basis of his genealogy. Whether he nevertheless believes in the existence of these drives

Greeks kept *the bad conscience* "at bay" by conceiving of their gods as taking the *guilt* upon themselves. In response to (2): it is true that Nietzsche says in GM II:24 that the bad conscience could be used in a different way from hitherto, but this does not in any way suggest that the latter should be distinguished from guilt. In response to (3): the point seems to already presuppose that bad conscience and guilt are distinct and that only the former but not the latter requires as a sufficient condition the internalization of cruelty, but no argument is made in support of this point. In my view, while it is correct that internalization of cruelty is not sufficient to explain guilt, it is not sufficient to explain bad conscience, either. All of this is not to deny that we can, if we wish, distinguish the two if we find it exegetically helpful (which I don't).

¹⁰ As Janaway comments, no such sharp distinction is operative in Rée's analysis, either, and it is plausible to assume, given Nietzsche's acquaintance with Rée's writings and the latter's interchangeable usage of these terms, that Nietzsche is more or less in agreement with Rée on this score (Janaway 2007, 129–130). Interestingly, Nietzsche's term "bad conscience"—*schlechtes Gewissen*—does not appear in Rée, in whom we can find the concept of the *evil* conscience—*böses Gewissen* (*Emergence*, section 9). Perhaps Nietzsche preferred the word "bad" here since the concept of "evil" serves a special role in his genealogy of slave morality in GM. Another term we can find in Rée is that of the "rebuking conscience"—*tadelndes Gewissen* (e.g., *Emergence*, section 7)—which does not occur in Nietzsche's text.

¹¹ Nietzsche is of the view that "the whole antithesis 'egoistic' 'unegoistic' obtruded itself more and more on the human conscience only when aristocratic value judgments *declined*" (GM I:2). It is only then that the word "good" began to be linked to the "unegoistic." As a consequence, one should not project this distinction into what one conceives of as lying at the basis of the genealogy of the moral term "good." This is usually taken to be another shot directed at Rée. But in all fairness, Rée, too, does not assume that "good" was right from the start associated with the unegoistic. As we saw in chapter 4, "good" in his view was initially used to refer to the useful, regardless of whether it is egoistically or unegoistically motivated.

is hard to say. There is evidence that at least at the time of HH, he did in fact believe that there is operative in us the “unegoistic drive [*unegoistischen Triebes*]” (HH 49)—a possible reference to Rée’s concept—but this drive seems to disappear from Nietzsche’s thought in the following years, and no similar drive or instinct is mentioned in the *Genealogy*. This is not to deny, of course, that Nietzsche conducts his psychological analyses in terms of the concept of drive. Quite the contrary: the concept of the drive, as well as that of “instinct,” which Nietzsche does not clearly distinguish from “drive” (e.g., GM II:16), is omnipresent in Nietzsche’s philosophy. In fact, Nietzsche is willing in a rather promiscuous manner to denote almost any form of mental activity, passion, or motivation with the designation “drive” (e.g., GM III:9).

Another related thing to note is the following. For Nietzsche, rather curiously, feeling guilty about hurting someone else (for instance) has nothing in essence to do with any natural tendency of human beings to care for each other. Rather, it is what Nietzsche calls the “instincts of freedom”—those of “Hostility, cruelty, joy in persecuting, in attacking, in change, in destruction” (GM II:16), which Nietzsche also groups together and refers to as “the will to power” (GM II:18)—that stand at the very core of the formation of the bad conscience and are presented as belonging to humans’ basic psychological endowment—Nietzsche’s psychological building blocks.

In what way, however, do these instincts form the basis of guilt or bad conscience? What *is* guilt or bad conscience? Here is the most basic characterization Nietzsche gives. The bad conscience or guilt in its “raw” or “animal” (GM III:20) form, involves (1) the inhibition (GM II:16, 17) of outward discharge of the “instinct of cruelty” (EH, Books, *Genealogy*), which Nietzsche conceives of as falling under the *active* “instincts of freedom” (GM II:16, 17, 18); (2) the consequent repression (GM II:17) or internalization (GM II:16) of this instinct; and (3) the discharge of this instinct “only on itself” (GM II:17) in a form of self-laceration (GM II:23) or

torture (GM II:6, 16, 22). “[T]hat”—Nietzsche sums up—“and that alone is what the *bad conscience* is in its beginnings” (GM II:17). As Nietzsche clarifies, the inhibition and internalization of cruelty are forced upon the subject by society. This is so because life in human society—becoming “enclosed within the walls of society and of peace”—requires the restraint of the external discharge of the cruel and hostile active forces, a restraint achieved by means of the “fearful bulwarks” society employs, chief among which is punishment (GM II:16). This is an element that Rée more or less presupposes in his discussion insofar as he posits punishment without much explanation of the social and psychological reality that stands at its basis and gives rise to it. Now, insofar as this process of forced internalization is something that takes place in every human society (cf. Ridley 1998, 21), we might expect the creation of the bad conscience in this sense to be omnipresent.¹²

Importantly, Nietzsche signals here with his talk of the “beginnings” of bad conscience that this internalization can become psychically entrenched at different times and places to different degrees and in different ways. Thus, at one point, Nietzsche describes the bad conscience as having become “firmly rooted” (GM II:21), which indicates that it can be more or less instilled, and that the manner in which it is interpreted, negotiated, and experienced varies from one epoch to another. Nevertheless, in *all* cases of bad conscience and guilt, we have a subject who discharges his internalized “instinct of freedom” inward at himself or herself (though there is an ambiguity here, as we will see). This minimal,

¹² There is, however, one curious exception: Nietzsche claims that those “blond beasts of prey” who set upon a population, laid its “terrible claws” upon it, and imposed a form on it—that is, prevented the members of the conquered population from freely discharging their instincts of freedom, thus creating the condition for the emergence of the bad conscience in them—“do not know what guilt, responsibility, or consideration are” (GM II:17). It is hard to imagine an organized group of people of whatever kind without at least *some* kind of self-restraint and hence *some* form of guilt or bad conscience. Perhaps Nietzsche’s idea is that these “beasts” demand so little of each other by way of inhibition that no significant internalization occurs to initiate the formation of a bad conscience.

“raw” sense of bad conscience is the common denominator in all cases. Crucially, though, mere internal discharge of repressed cruelty does not amount in itself to bad conscience or feelings of guilt—the discharge must be related in some way with the guilty agent’s recognition that he or she has done wrong in some sense.

A first question that arises at this point has to do with Nietzsche’s concept of repression or the “internalization of man.” Specifically, one may wonder why it is the case that drives or instincts that cannot discharge externally, that is, motivate the performance of actual deeds, are internalized. Here Nietzsche relies on a basic, general principle of his psychology according to which “All instincts that do not discharge themselves outwardly *turn inward*—this is what I call the *internalization* of man” (GM II:16). This has the psychological consequence that drives that motivate actions the performance of which is discouraged by means of fear, for example (GM II:15), do not, once their natural outlet is blocked, disappear but rather shift their direction and target not the external world but the internal world of the agent himself. They now “discharge” (GM II:17) themselves internally, rather than in public behavior. A crucial aspect of this internalization of the instincts of cruelty is that it gives rise, Nietzsche claims, to “what was later called” the “soul”: “The entire inner world, originally as thin as if it were stretched between two membranes, expanded and extended itself, acquired depth, breadth, and height, in the same measure as outward discharge was *inhibited*” (GM II:16). The more the discharge of cruelty was prohibited and the more cruelty was kept inside, the more the inner realm ballooned. But the more it ballooned, the more there grew an inner mental space—the “soul.” To extrapolate, what forms here is the capacity for self-consciousness and self-criticism. This condition of the soul being “at odds with itself” (GM II:18) is exploited at later stages by what Nietzsche calls the *active* bad conscience (GM II:18). What does Nietzsche have in mind with this latter concept?

Nietzsche, though admittedly not very clearly, makes an important distinction between two kinds of bad conscience—an *active*

and a *reactive* bad conscience—a distinction that is for the most part left undiscussed by commentators.¹³ To better understand this distinction, we need to explain briefly Nietzsche's distinction between active and reactive instincts or drives. While active instincts or drives arise in a subject spontaneously, that is, without being triggered or provoked by conscious events (an event the subject is conscious of) or external triggers (even if not given in consciousness), reactive drives or instincts appear only as a reaction, that is, only as a result of some conscious stimuli or external stimuli. This implies that humans' instincts of cruelty, which Nietzsche thinks of as active, arise and demand gratification independently of whether they are provoked by external conditions or by some experience or other. In contrast, *ressentiment*, for example, a central concept in Nietzsche's moral psychology, is reactive, for it *always* appears as a result of some instigation or other, specifically, as a result of displeasure of some kind or other; *ressentiment* never arises by itself.¹⁴

Now, what about the human being who is "enclosed within the walls of society and of peace" and cannot freely act on his active instincts of cruelty but internalizes them? Not only does this mean that these instincts will henceforth be discharged inward; it also means that they will become *reactive*. As Nietzsche puts it, in the form of the "feeling of guilt" or "bad conscience" or "the sting of conscience"—Nietzsche here makes it clear these are equivalent for him—they become a type of "psychical reaction [*seelischen Reaktion*]" (GM II:14; emphasis added). It is easy to see what Nietzsche has in mind here: the self-laceration that is characteristic of the *reactive* bad conscience is not spontaneous, that is, it does not erupt on its own but rather as a *reaction* to some offense or other that the agent recognizes that he has performed. The instincts of freedom, or at least a significant portion of them, then,

¹³ Though see Risse for a reference to what he terms "*locally-reactive guilt*" (Risse 2005, 36), which corresponds to the idea I will develop here.

¹⁴ For more on this, see Elgat 2017, chap. 2.

when “pushed back and repressed” (GM II:17), lose in the reactive bad conscience their active nature. The feeling of guilt is thus the becoming-reactive of cruelty.¹⁵

But what about the *active* bad conscience? Here Nietzsche gives us very little to work with, but the concept is introduced in section 18, where he elaborates a bit on the nature of the instinct of freedom (i.e., the will to power) and claims that it “creates for itself a bad conscience and builds negative ideals.” He then adds:

This secret self-ravishing, this artist’s cruelty, this delight in imposing a form upon oneself . . . and in burning a will, a critique, a contradiction, a contempt, a No into it, this uncanny, dreadfully joyous labor of a soul voluntarily at odds with itself [*mit sich selbst willig-zwiespältigen Seele*] . . . eventually this entire *active* “bad conscience” [aktivische “*schlechte Gewissen*”; emphasis in original] . . . the womb of all ideal and imaginative phenomena, also brought to light an abundance of strange new beauty and affirmation. . . . This hint will at least make less enigmatic the enigma of how contradictory concepts such as *selflessness*, *self-denial*, *self-sacrifice* can suggest an ideal, a kind of beauty. (GM II:18)

We can see here that internalized cruelty can also operate “voluntarily,” that is, actively, and impose a “contradiction” on the soul in the form of certain negative ideals of self-denial. Here, too, the internalized cruelty is vented inward, but unlike the bad conscience in its reactive form, it does not wait for an occasion to discharge itself upon the subject but takes the initiative, so to speak, by erecting ideals that encourage forms of behavior that are self-negating and thus provide the occasion for acts of self-cruelty. These ideals, I want to suggest, are those that Nietzsche discusses in the third essay under the heading “the ascetic ideal” in its priestly

¹⁵ The becoming-reactive of active forces is a theme in Deleuze’s reading of Nietzsche (Deleuze 2006).

form (which he also describes as “contradictory”; e.g., GM III:13). All of this, as Nietzsche writes, takes place only “eventually,” at a much later stage. The bad conscience in its raw state, in contrast, first takes shape in “those millennia before the history of man” (GM II:14). In what follows, I will mostly focus on the reactive bad conscience or guilt.¹⁶

We should ask, however, how does it come about that one’s transgressions trigger the reaction of the feeling of guilt? How does a link form between such infractions and the unleashing upon oneself of one’s repressed cruelty? Without an answer to this crucial question, Nietzsche’s account would remain embarrassingly incomplete.

I wish now to flag an ambiguity in Nietzsche’s characterization of bad conscience. On the one hand, as we saw, Nietzsche claims that the instincts of freedom turn against *themselves* (*an sich selbst*) (GM II:17) in “a declaration of war against the old instincts upon which [their possessor’s] strength, joy, and terribleness had rested hitherto” (GM II:16). On the other hand, Nietzsche claims that in bad conscience, the instincts of freedom are “turned against the possessors [*Inhaber*] of such instincts” (GM II:16; see also Nietzsche’s remarks about the animal lacerating *itself* in GM II:23). So which one is it, and how to account for this apparent indecisiveness? It seems Nietzsche wishes to retain both of these separate, though related, features of bad conscience. One idea is that the negative feeling of bad conscience is experienced as directed at *oneself*, where this should be distinguished from its being about one’s *worth* as a person or one’s character or being. On the other hand, what one feels bad *about* is one’s actions; specifically, one feels bad *for* acting on one’s “instincts of freedom.” Putting the two together, we get the notion that one condemns oneself *for* performing such actions.

¹⁶ I am thus of the view that Reginster’s reading of GM II:18 is off the mark, chiefly because he does not address the crucial qualification of the bad conscience as “active.” In his reading, the “negative ideals” are formed by internalized *ressentiment*, but this is the reactive force par excellence! (Reginster 2018, 13–14).

If this is right, then bad conscience is essentially concerned for Nietzsche only with those actions that are motivated precisely by the internalized instincts of “Hostility, cruelty, joy in persecuting, in attacking, in change, in destruction.”¹⁷

This raises the question of how the “masters” of the various periods could act violently toward others (slaves, barbarians) without suffering from bad conscience: why don’t they feel guilty for their atrocious behavior (GM I:11)? The answer seems to be that not *all* forms of external discharge of the instincts of freedom are prohibited and thus restrained in any given community at every stage of civilization. Which types of hostile or cruel actions are allowed and which are prohibited depends on the nature of the community, that is, on the content of the “few primitive demands of social existence” (GM II:3), on the *morality of custom* (*Sittlichkeit der Sitte*) (e.g., D 9) that constitutes—along with its various punitive practices—the “social straight jacket” (GM II:2) in which one finds oneself confined. Moreover, as AOM 89 suggests, customs do not apply equally to those who are at the top of the social hierarchy and those at the bottom, to masters and commoners. Consequently, not everyone in the community is forced to internalize in the same way and cultivate a similar ability to experience guilt.¹⁸

IV. Prehistoric Stage and the Morality of Custom

What, then, is the morality of custom? The morality of custom, the prehistorical stage where the bad conscience is taking its first steps,

¹⁷ Nietzsche thus seems to be rather tame in his genealogy compared to Rée’s later *Emergence*, for in the latter, as we saw in chapter 4, Rée claims on the basis of anthropological data that in some societies, it is actually *violence* that is praised and *compassion* that is condemned, so that bad conscience attaches to the latter!

¹⁸ See Morrisson 2003, 665–666. In the following discussion of the morality of custom, I was helped by Morrisson’s paper.

consists of a set of customary practices that govern a community and determine the realm of what is considered ethical (*sittlich*) in that community. The “chief proposition” of the morality of custom declares: “morality [*Sittlichkeit*] is nothing other (therefore no more!) than obedience to customs [*Sitten*]” (D 9). How does the morality of custom originate?

Here we should distinguish a number of stages. The first involves the “ineluctable disaster” when

some blond pack of blond beasts of prey, a conqueror and a master race which, organized for war and with the ability to organize, unhesitatingly lays its terrible claws upon a populace perhaps tremendously superior in numbers but still formless and nomad. That is after all how the “state” began on earth. (GM II:17)

It is this “master race” which “subjugates all other individuals [*Einzelnen*], that is to say draws them out of their isolation and orders them within a collective” (HH 99). It is this act of shape giving that compels the subjugated with the threat of *punishment* to abide by certain *rules of conduct* and at the same time hinders the subjugated from discharging their “instincts of freedom,” which forms the basis for the bad conscience. It is thus by means of violent compulsion that the individual is integrated—socialized—into the morality of custom.

But we should now ask how the customs instituted by the conquerors attain the particular content that they possess. Why do they involve the specific rules they do and not others? Nietzsche claims that

How the tradition has *arisen* is here a matter of indifference, and has in any event nothing to do with good and evil or with any kind of immanent categorical imperative; it is above all directed at the preservation of a *community*. (HH 96; cf. BGE 201)

Despite his (later) objections to the utilitarian perspective, Nietzsche's view here is that the original customs are adopted because they are believed to be useful for the preservation of the community, not because they are taken to have any value in and of themselves or to have some kind of unconditional authority like the Kantian categorical imperative. As Nietzsche explains in the same section, even "some chance event mistakenly interpreted" or "superstition" can coalesce into custom (HH 96). This perspective of utility is expressed, Nietzsche explains, by the following twin principles: "the community is worth more than the individual [*Einzelne*]," and "an enduring advantage is to be preferred to a transient one," "from which it follows," Nietzsche claims,

that the enduring advantage of the community is to take unconditional precedence over the advantage of the individual [*Einzelnen*], especially over his momentary wellbeing but also over his enduring advantage and even over his survival. (AOM 89)¹⁹

Specific actions are thus weighed with respect to their being in accordance with custom—with what helps preserve the community; the egoistic/unegoistic distinction is "not the fundamental antithesis" (HH 96).²⁰ In time, "as a result of a long inheritance" (HH 96), these customs become *tradition* (*Herkommen*). This transition involves a change in the nature of the authority of the ethics of custom: one comes to obey customs *because* they constitute tradition. And what is tradition? Tradition is "A higher authority which

¹⁹ As Nietzsche clarifies in the same section, these principles apply to those at the bottom, the "sacrificial beast"; those at the top, the conquerors, rate "present enjoyment" "higher than an insipid living-on in a painless condition of comfort" (AOM 89). See Morrisson (2003, 665–666) for some discussion, and see also Nietzsche's claim that "whoever wanted to elevate himself above [the ethics of custom] had to become a lawgiver and medicine man and a kind of demi-god; that is to say, he had to *make customs*" (D 9).

²⁰ Rée, at least at the time of *Origin*, would, of course, concur: utility is the basic germ; focus on motivation appears later.

one obeys, not because it commands what is *useful* to us, but because it *commands*” (D 9). This, however, does not mean that the utilitarian perspective is completely abandoned; it is retained, but under a different guise: one obeys out of fear and respect for the gods. To prepare for the introduction of this idea, we have to take a few steps back and examine a couple of parallel lines of development that Nietzsche explores in the *Genealogy*.

Toward the beginning of the genealogy of guilt, Nietzsche makes a hypothesis, taking his cue from German etymology, that “the major moral concept guilt [*Schuld*] has its origin in the very material concept debts [*Schulden*]” (GM II:4). And as he puts it later, “the feeling of guilt [*das Gefühl der Schuld*], of personal obligation, had its origin, as we saw, in the oldest and most primitive personal relationship, that between buyer and seller, creditor and debtor” (GM II:8).²¹ In what sense does the origin of the feeling of guilt lie in the creditor-debtor relationship?

After criticizing the “the genealogists of morals” (GM II:4) for their lack of historical insight and for thinking that guilt and punishment were right from the beginning tied to the idea of free will and the ability to act otherwise, Nietzsche goes on to claim that punishment “throughout the greater part of human history” was not administered on the basis of the ideas of responsibility or freedom. Rather, punishment was an expression of anger, an anger that was nevertheless “held in check and modified by the idea that every injury [*Schaden*] has its *equivalent* and can actually be paid back, even if only through the *pain* of the culprit.” And Nietzsche

²¹ As is well recognized in Nietzsche scholarship, the ambiguity of *Schuld* in German multiplies the number of possible interpretations of the text, for almost every instance of the German *Schuld* raises the question of whether it should be read as “guilt” or “debt.” Here, for example, in GM II:8, in agreement with translators Kaufmann and Diethel, I think the correct translation is “feeling of guilt” rather than “feeling of debt,” because Nietzsche seems to repeat here the idea he presents initially in GM II:4, where the *moral* concept is claimed to have its origin in the material concept of debt. Further, “feeling of debt” sounds unnatural in English, while the more precise “feeling of indebtedness” would be *Gefühl der Verschuldung* in German, and not *das Gefühl der Schuld*, as we have it in the text.

immediately asks, “And whence did this primeval, deeply rooted, perhaps by now ineradicable idea draw its power—this idea of an equivalence between injury and pain?” and answers, “I have already divulged it: in the contractual relationship between *creditor* and *debtor*, which is as old as the idea of “legal subjects” and in turn points back to the fundamental forms of buying, selling, barter, trade, and traffic” (GM II:4). Nietzsche thus states rather clearly what importance the creditor-debtor relationship has for him: *it is here that the equivalence between injury and pain first takes form*.²²

Why, however, would a relation of equivalence between injury and pain form in the creditor-debtor relationship? The idea seems to be, as Nietzsche’s short foray into the primeval joys of cruelty in punishment and in general suggests (GM II:5–6), that a debtor who “injured” the creditor, that is, who *reneged* on his debt, could then expect to be visited upon by “every kind of indignity and torture” for which “everywhere and from early on one had exact evaluations” “commensurate with the size of the debt” (GM II:5)—*this* is the equivalence between injury and pain that Nietzsche has in mind. The mere *having* of a debt, by itself, then, was not associated with *any* kind of punishment or pain. It is only the *failure* to repay the debt that had pain follow in its train as punishment. Of course, since there are no fixed and precise standards by means of which one can determine precisely how to equate injury with punishment, what is taken as “equivalent” will differ from one societal context to another (see GM II:10).

²² Reginster argues for a greater importance of the creditor-debtor relationship for an understanding of the nature of guilt. According to him, focusing on this relationship uncovers a feature concealed by the usual conception of guilt as responsibility for a moral transgression, namely, that guilt involves a feeling of indebtedness, of *owing* something (Reginster 2018, 3, 10). While an insightful suggestion, I don’t think it is supported by the text. Rather, as we saw, for Nietzsche the importance of the relationship lies in the formation of the link between injury and pain or punishment. Perhaps more important, guilt, in my view, does not necessarily involve the feeling that one owes something to someone: I can feel guilty for running my car over a tortoise without feeling that I am now indebted to anyone or anything in any way. But neither is this feeling sufficient, for it is not an exclusive characteristic of guilt; thus, the recipient of a favor might feel indebted to the benefactor, sans guilt.

At this early stage, it should be emphasized, debtors are not sufficiently skilled in adhering to their side of the deal: they are indeed expected to pay the debt back, but cannot at this early stage take their “obligations” (GM II:6) seriously enough²³ and do not yet possess the personal capacities to do so, capacities I will discuss when I turn to examine the “sovereign individual.” In other words, debtors at this stage are still not sufficiently *responsible*.

In the next stage of the genealogy, Nietzsche presents us with the *first* interpretation and extension of the creditor-debtor relation whereby it is broadened beyond its original sphere and transposed onto the relation in which the community stands to its single members (though he unfortunately does not elaborate on the nature of this transition). As a result of this interpretation, “the community, too, stands to its members in that same vital relation, that of the creditor to his debtors” (GM II:9). If an individual member (*Einzelnen*, GM II:10) fails to abide by the customs of society—if one *injures* the community—one gets punished by the community, the “disappointed creditor” (GM II:9). Indeed, the “lawbreaker is a debtor who has not merely failed to make good the advantages and advance payments bestowed upon him but has actually attacked his creditor” (GM II:9). Infractions of custom are perceived as hurting the community, since, as we saw, custom is taken to be necessary for its preservation. As Nietzsche puts it in HH 96: “to sever one-self from [the community] is dangerous, and even more injurious to the *community* than to the individual [*Einzelnen*].” Nietzsche is imagining here a rather primitive form of collective human existence but one where there is nevertheless some agent of authority—those founders of the community, the “beasts of prey,” who possess the power to punish those who go against the customs.

²³ To take the obligation seriously is, at a basic level, to remember it and try one’s best to discharge it out of fear of punishment; at a much higher level and a later stage, to commit oneself is to regard one’s obligation as *intrinsically* authoritative—one should keep one’s obligations because this has value in itself.

What are the *effects* of punishment in this earliest of times? As we shall see, the effects of punishment—especially the *pain* it causes and the *fear* it induces—are prodigious, and yet Nietzsche is careful to emphasize that at this most primitive stage, punishment does not have the effect of arousing the “sting of conscience”; quite the contrary, its development, through punishment, “was most powerfully *hindered*—at least in the victims upon whom the punitive force was vented” (GM II:14). The process of the forming of bad conscience is at this stage already under way—discharge of cruelty is already inhibited—but is, paradoxically, held back by precisely those forces which created the inhibition in the first place. This is because, Nietzsche presumes, once the culprit comes to witness and experience firsthand the “judicial and executive procedures” employed by the agents of authority in the community and sees how these procedures are practiced by them “with a good conscience”—that is, without any compunction—he will *not* come to “feel [*empfinden*]” that such a “kind of actions”—“spying, deception, bribery . . . robbery, violence . . . torture, murder”—*including his own*, are “reprehensible [*verwerflich*]” “as such [*an sich*]”: such forms of behavior are only condemned by the culprit’s judges “when they are applied and directed to certain particular ends” (), that is, when they are seen as putting the community at risk (GM II:14). Nietzsche’s implicit assumption is thus that to consider the action as reprehensible “as such” is a necessary condition for the experience of bad conscience. In other words, it turns out that punishment and the internalization of the instincts of cruelty, though necessary, are not sufficient for bad conscience. The criminal does indeed direct some internalized cruelty against himself in the form of accusations of imprudence or carelessness and berates his own instincts for getting him in trouble—the realization that certain courses of action are punishable is, after all, what the internalization of the community’s norms involves—but he does not criticize his deed as such. Importantly, in addition, punishment at this stage

does *not* give rise to the morally charged thought “I ought not to have done that” [*das hätte ich nicht thun sollen*]” (GM II:15).

What is it, exactly, to consider certain actions wrong *as such*? GM II:14 suggests two answers. The first is that to consider an action wrong as such is to think that *no matter the consequences*, one should avoid performing it. The second is that to consider an action wrong as such is to regard the action as reprehensible *no matter who* performs it—king or beggar, master or slave, the culprit or the “judges.” These two possibilities are distinct: the same kind of action can be regarded wrong as such in one sense but not in the other. Thus, an action can be regarded by a noble person as wrong as such in the sense that it is to be shunned not because of its bad consequences but because it is improper to be performed (only) by a person belonging to the noble caste. Crucially, in the case of a *moral* evaluation of an action (at least of the Kantian kind, at issue also for Rée), the two come together: an action is morally wrong when it is wrong *no matter who* performs it and regardless of the consequences. To explain moral guilt, therefore, we would have to explain how an action can come to be regarded as reprehensible as such in both senses.

In Rée’s genealogy of morality, as we saw, certain actions, specifically, those of the egoistic kind, are initially not condemned *as such* but are criticized because of their perceived detrimental consequences to the well-being of the community. The same idea seems to be operative here, in the first stages of Nietzsche’s genealogy. In Rée’s genealogy, however, in the last account, it is only as a result of human forgetfulness that egoistic actions come to be negatively valued as such. Does Nietzsche appeal to forgetfulness in his account as well? How else can he explain the disvaluation of a kind of action “as such,” crucial to the experience of moral guilt? Without providing an answer, Nietzsche’s account, it seems, threatens to leave us with a gaping hole in our understanding of the genealogy of guilt.

To return to the question of the effects of punishment: as a result of the socio-psychological practice of punishing the criminal “debtor,” an “uncanny intertwining [*Verhäkkelung*] of the ideas ‘debt [*Schuld*] and suffering’ was first effected—and by now they may well be inseparable” (GM II:6).²⁴ What does Nietzsche have in mind here? Unfortunately, he does not elaborate on what I think is an absolutely crucial thread in the cloth he is weaving. The idea, I think, concerns a psychological development whereby the idea of having a debt, or rather, to be more precise, the idea of *failing* to pay a debt, that is, failing to abide by the norms that govern a community, comes to be *associated* in the individual with the idea of pain. Human subjects who participate in the creditor-debtor relation (with respect to others or the community) and who get punished for their transgressions come to associate psychologically failure to pay back their debts with pain, with the result that transgressions give rise in the mind of the criminal to anticipatory feelings of pain. A similar connection can form in the minds of observers. And given that one of the first effects of punishment was an “increase of fear” (GM II:15), it would not be a stretch to hold that this anticipation of suffering is colored by a feeling of fear of the punishment that one believes would descend upon one’s head “like a piece of fate” (GM II:14). With this, we have something close to an explanation of the feeling of guilt: a transgression gives rise to an unpleasant feeling of fear of an impending punishment. For all of Nietzsche’s animosity toward the English psychologists, with his talk of “intertwinement,” Nietzsche seems to have something very close to the associative psychology we found in Rée’s genealogy.

²⁴ I agree with Diethel’s translation—in contrast to Kaufmann’s—that *Schuld* should be translated here as “debt”; this makes more sense, because “guilt” already involves in itself a form of suffering. Moreover, at this stage of the genealogy, we have not yet been given an answer regarding how guilt forms, so it would make little sense for Nietzsche to claim that the punitive practices of the primeval communities brought about an “intertwinement” of guilt with suffering.

This, however, would still not be sufficient as an account of guilt for at least a couple of reasons. First, as we saw with Rée, fear should be distinguished from moral guilt proper, and here we seem to have mainly the former. Second, this account says nothing about the role of internalized cruelty in guilt—central in Nietzsche’s analysis. It is possible, of course, to hold that the culprit, when caught, will not only fear her punishment but will also unleash her cruel instincts upon herself. While correct in principle, we are missing a psychological explanation of *how* precisely the unleashing of cruelty upon oneself is triggered by the realization that one has violated the community’s mores.

To make further progress in our genealogy, we need to introduce a *second* interpretation, a second expansion that the debtor-creditor relationship undergoes. As Nietzsche claims,

this relationship was for a second time transformed through interpretation, in a historically extremely strange and curious manner, into a relationship in which it is perhaps least comprehensible to us modern men: that is the relationship of the *present generation* to their *forebears*. (GM II:19)²⁵

The idea is that every generation in the community comes to feel indebted to its ancestors, especially to those who “founded the tribe,” for it is thanks to the achievements of those “that the tribe *exists*”: “one thus recognizes a *debt* that constantly grows greater, since these forebears never cease . . . to accord the tribe new advantages and new strength.” How does one pay them back? “Sacrifices . . . feasts, music, honors; above all, obedience—for all customs, as works of the ancestors, are also their statutes and commands.” A fateful transformation occurs as time passes by: in the consciousness of the living generation, the ancestors turn into

²⁵ I use Diethe’s translation in this quotation; Kaufmann leaves out the crucial “for a second time” from his translation.

“powerful spirits” and gods (GM II:19). Nietzsche gives two complementary accounts of how this transformation of human into divine occurs. In the earlier HH 96, he claims that

Every tradition now continually grows more venerable the farther away its origin lies and the more this origin is forgotten; the respect paid to it increases from generation to generation, the tradition at last becomes holy and evokes awe and reverence [*Ehrfurcht*].

Here the status of the traditional customs themselves undergoes transformation, and the laws of the community become divine in part by virtue of (Rée-ian) forgetting. This transformation accounts for the change in the authority of the customs. As these become tradition, and as the origin of the tradition recedes farther and farther into the past, the customs become sacred and attain a validity that transcends, to some extent, the utilitarian: one obeys tradition out of “reverence”—a mixture of fear and respect (see also D 9). One now obeys *partly* because of the categorical nature of the customs and *partly* because of the fear of the horrible consequences that will befall the community if customs are not followed. In other words, one no longer obeys customs on the basis of the belief that obedience is a *direct* means to the preservation of the community. Rather, one obeys because the customs attain the status of something like a *taboo*: they attain a mysterious and unconditional authority, but one at the same time realizes that if they are *not* followed, some grave disaster will strike. It is perhaps here that agents for the first time begin to see that their injurious actions also have an aspect to them by virtue of which they are wrong *as such*. Consequently, their self-laceration acquires, to some extent, the nature of bad conscience. Importantly, this is not yet *moral* guilt, for here bad conscience is still qualified by fear, and the violations that give rise to it are violations of the dictates of some external authority, which we

might also regard (in Kantian fashion) as not belonging to morality in the strict sense.

In the *Genealogy*, the account of the transformation of the human ancestors to gods focuses more on the becoming-divine of the ancestors themselves, rather than their laws:

the ancestors of the *most powerful* tribes are bound eventually to grow to monstrous dimensions through the imagination of growing fear and to recede into the darkness of the divinely uncanny and unimaginable: in the end the ancestor must necessarily be transfigured into a *god*. Perhaps this is even the origin of gods, an origin therefore out of *fear*! (GM II:19)

The idea seems to be that since the power of the community is regarded as a result of the actions of its founding members, the stronger and more prosperous the community grows, the more powerful the founding members become in the imagination of the community's members. When this logic is pushed to its extreme, the founding members appear to be godlike, divine. Though in HH, Nietzsche still relies on the psychological effects of forgetfulness to account for the becoming-holy of tradition (see also WS 40), here, in the *Genealogy*, as part of his attempt to distance himself from "the English genealogists," Nietzsche avoids its explicit invocation. Nevertheless, it is implicitly present in the *Genealogy* account as well, for without forgetting that the ancestors were human, all too human, the imagination could arguably not carry out its work of inflating the mortal into the divine. Thus, forgetfulness à la Rée is necessary to transcend the level of mere fear as motivation for obedience.

As a result of this double expansion of the creditor-debtor relationship, every member of the community owes his or her compliance *both* to society (GM II:9) *and* to the gods (GM II:19); debt governs both interpersonal relationships and the human-divine relationship. Thus, all new members of the community inherit debts,

and “this inheritance then overflows them in all directions” (GM II:20). Correspondingly, when a member of the community fails to abide by custom (fails to repay the debt), he or she could expect to be punished by *both* creditors. But there is a difference: while society can isolate the culprit and vent its wrath upon him or her alone, divine wrath befalls the community *as a whole*. As Nietzsche claims, “supernatural punishment,” “punishment for breaches of custom will fall before all on the community” (D 9); “the gods punish the community for misdeeds and for every violation of their privileges and only to that extent punish the individual” (HH 96). On the other hand and at the same time,

The community can compel the individual [*Einzelnen*] to compensate another individual of the community for the immediate injury his action has brought in its train; it can also take a kind of revenge on the individual for having, as a supposed after-effect of this action, caused the clouds and storms of divine anger to have gathered over the community. (D 9)

V. The Sovereign Individual

The genealogical account given so far leaves us with an explanation of how a member of a community of the morality of custom can come to feel a very basic form of bad conscience: fear upon the performance of actions that go against the dictates of customs, a fear, however, that possesses a particular hue insofar as it arises out of the recognition that one violated the sacred commands of the gods. A number of things should be noted. First, we still have no answer to the question regarding the role of internalized cruelty in the formation of this quasi-guilt. Second, as Nietzsche himself emphasizes (GM II:4, 14, 15), at this stage, there is neither awareness of personal responsibility for the action nor is the misdeed accompanied by the thought of free will (the thought “I could have acted differently”).

At this stage, there is still no developed capacity to experience *individualized* bad conscience/guilt. Third, we still lack an explanation for how *moral* guilt forms, the unpleasant feeling for doing something wrong *as such*.

To address some of these outstanding issues, we have to turn to the psychological type Nietzsche calls “the sovereign individual [*das souveraine Individuum*]” (GM II:2). The figure of the sovereign individual has for more than a decade stood at the center of an exegetical debate concerning its identity and ideality in Nietzsche’s eyes. On the one hand, there are those who hold that the sovereign individual represents for Nietzsche some kind of ideal (Ansell-Pearson 1990, Zamosc 2012, Reginster 2018) or embodies his “positive view of freedom” (Richardson 2009, 128). On the other hand, there are those who think that while the sovereign individual represents “the modernist ideal of subjective autonomy,” this ideal is one which “Nietzsche displaces” (Hatab 1995, 37; see also Acampora 2006).²⁶ Expressing a similar, critical view is Leiter (2011), according to whom Nietzsche, by presenting the sovereign individual as if it represented a significant, world-historical achievement, is merely employing ironic rhetoric to mock the domesticated, human animal which has been trained to perform the trick of making and keeping promises.²⁷

What is often lost sight of in these debates, however, is the role of the figure of the sovereign individual in Nietzsche’s genealogy of guilt and bad conscience.²⁸ On my reading, Nietzsche’s account of

²⁶ But, pace Hatab, the autonomy at issue in the case of the sovereign individual, I wish to argue, is not to be traced back to the free subject of GM I:13. The autonomy of the sovereign individual—and here I am in agreement with Gemes (2009) and Reginster (2018)—is to be understood in a compatibilistic sense. I will return to this.

²⁷ To be fair, Leiter also provides a second “deflationary reading” where the sovereign individual *does* represent an ideal, “but such a self, and its self-mastery is, in Nietzschean terms, a fortuitous natural artifact (a bit of “fate”), not an autonomous achievement for which anyone could be responsible” (Leiter 2011, 103). Leiter tends to think that the deflationary reading is “probably the correct one” but finds the first reading, where the figure of the sovereign individual is “wholly ironic,” attractive as well.

²⁸ Zamosc (2012) and Reginster (2018) are notable exceptions. I will address some of the points they make.

the emergence of the sovereign individual is an indispensable element in Nietzsche's genealogy of bad conscience and guilt.

In this section, I will argue for the following claims. First, that the figure of the sovereign individual is not a singular or unique occurrence in Nietzsche's published writings but is clearly present in certain sections from *Daybreak* and *The Gay Science* that have not to my knowledge been pursued in the anglophone literature. Second, that examination of these texts (especially D 9) reveals that the sovereign individual is for Nietzsche embodied (though not exclusively) by Socrates and the Socratic individual. Third, and most important, that at least part of what is novel about the sovereign individual is that it is only with the sovereign individual that there emerge *personal responsibility* and *individualized guilt*—the self-laceration that concerns one's misdeeds *alone* and involves the belief that one's misdeeds are entirely one's *own* individual fault, rather than the fault of the community or the gods—essential components in our sense of guilt. I start with an analysis—conceptual and genealogical—of some of the capacities that the sovereign individual possesses. Understanding these capacities will enable us to grasp better the sovereign individual's emergence and identity.

a. The Production of the Sovereign Individual: The Origin of Responsibility

Nietzsche opens the second essay of the *Genealogy* with the question of how it is possible to “breed an animal *with the right to make promises*” (GM II:1), but—and here I am sympathetic to Zamosc's (2012) and Reginster's (2018) approaches—Nietzsche is actually after something more substantial of which the right to make promises is only a paramount instance: the “breeding” of a responsible agent. Nietzsche thus first enumerates capacities whose gradual acquisition makes up “the long story of how *responsibility* originated” and then presents the “*sovereign individual*” as the

“fruit” of the “tremendous labor of [the] ‘morality of mores’” (GM II:2). This, as well as how the sovereign individual is subsequently described, strongly suggests that the sovereign individual presents us with a highly developed instance of agency, where responsibility reaches a certain high point. This sovereign individual, “liberated again from the morality of custom,” moreover, possesses “power over oneself and over fate,” enjoys—as a result, we can assume—a high level of “pride” in his response-*ability*, and is “autonomous” (GM II:2). But how does the “tremendous labor of [the] ‘morality of mores’” produce this supremely responsible agent? We have already encountered some of the effects that punishment can have on the members of the community. I have, however, so far omitted from discussion the effect of punishment on the production of several capacities necessary for the creation of the more or less reliable members of those early communities; at their height, I will argue, these capacities present us with the sovereign individual.

I start with the production of memory. Nietzsche first explains that to counteract humans’ natural forgetfulness, a genuine “*memory of the will*” is required, namely,

an active *desire* not to rid oneself, a desire for the continuance of something desired once . . . so that between the original “I will,” “I shall do this” and the actual discharge of the will, its *act*, a world of strange new things, circumstances, even acts of will may be interposed without breaking this long chain of will. (GM II:1)

In other words, to have a real memory of the will requires the will and the ability to “stick to it,” to keep your eyes on the ball, so to speak, without being distracted away from it, until you can finally discharge your obligation. More concretely, one would not be able to abide continuously and reliably by the laws of the community if one could not remember that and how these laws applied to oneself or if one was easily distracted so that one’s volitional “chain” was easily broken. Without keeping the community’s laws in mind and

acting accordingly, one would behave like an irresponsible child. This, the creation of memory, Nietzsche immediately explains, requires a set of capacities ("how many things this presupposes!"; GM II:1) that, as he clarifies, stand at the origin of responsibility (GM II:2). What are these capacities? To be responsible,

man must first have learned to distinguish necessary events from chance ones, to think causally, to see and anticipate distant eventualities as if they belonged to the present, to decide with certainty what is the goal and what the means to it, and in general be able to calculate and compute. Man himself must first of all have become *calculable*, *regular*, *necessary*, even in his own image of himself, if he is to be able to stand security for his own future, which is what one who promises does! (GM II:1)

A bit later in the next section, Nietzsche clarifies that the production of responsibility and thus of sovereignty also "presupposes as a preparatory task that one first *makes* men to a certain degree necessary, uniform, like among like, regular, and consequently calculable," for which "the social straightjacket" of the "morality of mores" is expedient (GM II:2). The central idea that emerges here is that the production of a responsible agent presupposes the acquisition of the interrelated traits of (1) being *calculable* (*berechenbar*), and (2) being able *to calculate* (*rechnen*).

First, what is it to be calculable? An agent is calculable if others can "calculate" his or her actions, that is, if they can predict with some level of assurance what this agent will do in certain circumstances. This requires, among other things, knowledge of what the person values. Enforcing certain values on the members of a community is precisely the morality of custom's way of introducing a high level of regularity that makes such predictions possible: one will be able to predict the agent's behavior since society has trained him or her, by means of certain painful "instruments of culture," to retain in the mind—to remember—"a few primitive demands of social existence

as *present realities*" (GM II:3). For a social demand or law to be present in one's mind as a present reality is for it to be efficacious in producing the agent's actions in a reliable way. In this way, the community's values become, as Nietzsche puts it elsewhere, "the most powerful levers in the involved mechanism of our actions" (GS 335).

The memory of these "demands," however, is not sufficient, for the agent must also be able to overcome momentary temptations and distractions that can have the effect of pulling him or her away from abiding by these laws. In other words, to overcome such opposing forces, the agent requires *self-control*. "To see . . . distant eventualities as if they belonged to the present" (GM II:1) is part of this capacity. A person who discounts the future in favor of the present (thus failing to see it as if it belonged to the present) will be tempted by present attractions, since the future consequences of succumbing to such attractions will appear distant and thus insignificant in comparison to present advantages. In contrast, an individual who knows what the laws of the community are, insofar as these have been *burned* by means of punishment into his or her memory, and is, moreover, *able* to abide by these laws will tend to act in conformity with them. It is precisely by virtue of this that the ultimate product of these processes, the sovereign individual, is regarded as *reliable* (*Zuverlässigen*) (GM II:2).

More, however, is needed. As Nietzsche explains, to become responsible, one must also possess the ability to *calculate*, that is, think practically. All the self-control in the world won't help an agent successfully carry out a task that is expected of him or her without an ability to *plan* ahead, "to ordain the future in advance" (GM II:1). It is easy to see how the abilities Nietzsche is here singling out—the ability to think causally, to distinguish "necessary events from chance ones," and to decide "what is the goal and what the means to it" (GM II:1)—are immediately relevant, for they lie at the very foundation of competent agency. Without telling the necessary from the contingent, an agent can't know which aspects

of reality he or she can manipulate, and without familiarity with causal relations, an agent can't tell how this manipulation is going to result, specifically, whether it will enable him or her to achieve the goal or not.

Finally, the agent must also be seen as calculable "in his own image." In other words, the agent must also possess a form of self-awareness, for without believing that he can keep his promises, the agent won't be justified in his promise making—he won't possess the right to make promises. How are all these capacities acquired?²⁹

Here we have to revisit GM II:14, where, as we saw, Nietzsche avers that at the beginning, punishment does not arouse the "feeling of guilt" (GM II:14): rather, "it was prudence [*Klugheit*] that criticized the deed" (GM II:15). In other words, the criminal criticized himself for not being careful enough or clever enough. In contrast to the "psychic reaction of 'bad conscience'" (GM II:14), Nietzsche explains, "the actual *effect* of punishment must beyond question be sought above all in a heightening of prudence, in an extending of the memory . . . in a kind of improvement in self-criticism . . . mastery of the desires" (GM II:15). Furthermore, Nietzsche clarifies that the transition to society involved the forcing of the "semi-animal," the human, to "thinking, inferring, reckoning, co-ordinating cause and effect" (GM II:16). We thus see how punishment, that is, the inflicting of *pain*, serves as a central means for the creation of some of the elements that gradually produce full-fledged responsibility.

But how can pain be of help in the creation of the capacities just mentioned? With the forming of the "intertwinement" mentioned earlier, the agent begins to connect ethical violations with pain and begins to see that the one is the cause of the other. In other words, the infliction of pain for transgression is in itself a lesson in causal thinking: it makes it painfully evident that certain actions have consequences. In addition, in GS 13, Nietzsche claims that in

²⁹ Assuming an internalist conception of justification.

contrast to pleasure, “pain always raises the question about its origin.” In other words, pain cultivates backward causal thinking as well in that pain prompts the questions: How did this happen that I am now punished? Where did I go wrong? What did I fail to see? And a result of such self-interrogations is a sharpening of prudence (causal thinking and planning) and a heightening of self-control. In this way, punishment gradually hones precisely those abilities that help bring forth responsibility, and the more these capacities develop, the more progress is made toward the sovereign individual.

How does the sovereign individual differ, then, from a pre-sovereign subject of the morality of custom? One answer is that the sovereign individual, in contrast to the person who has not yet *fully* learned to calculate and be calculable, is genuinely responsible and thus possesses the “right to make promises.” While correct, this answer does not explain the role of the sovereign individual in the genealogy of the bad conscience and guilt. Importantly, it cannot be that it is only with the sovereign individual that the ability to feel guilt is produced, for according to Nietzsche—and I will return to this—persons living under the morality of custom could already experience guilt in some sense.

Is it possible, as Reginster (2018) has suggested, that it is only with the sovereign individual that the ability to feel *moral* guilt emerges? Reginster rejects the interpretation that the sovereign individual is a moral nonconformist who creates his own values (2018, 5). Rather, the sovereign individual, according to Reginster, retains the values that inform the morality of custom, his “five or six ‘I will not’s” (GM II:3). Where he differs is with respect to the structure of his motivation: “the sovereign individual is liberated from the morality of custom insofar as the motivation of his promise keeping is no longer the fear of punishment” (Reginster 2018, 5). The sovereign individual “cares about keeping his promises “as such”” (6) and is motivated to keep his promises—and by extension, his adherence to the values of his community—not by fear for “his welfare” (5), which motivates the subject of the morality

of custom, but by a sense of pride: his worth as a person depends on living up to the commitments he has taken upon himself, and it is this non-prudential motivation that (partly) accounts for moral guilt.³⁰

Though I am sympathetic to the notion that with the sovereign individual, a change in the structure of motivation occurs, there are at least three problems with Reginster's view. First, the subject of the morality of custom's motivation to adhere to the ethics of custom is not *simply* one of fear. As we saw, it is fear intermixed with respect, or reverence, that is the ruling motivation here: what "distinguishes this feeling in the presence of tradition from the feeling of fear in general" is that it is fear "in the presence of a higher intellect which here commands, of an incomprehensible, indefinite power . . . there is superstition in this fear" (D 9). And it is this incomprehensibility that undergirds the "*respect* [*Achtung*] for the authority of custom" (D 10; emphasis added). Second, Reginster is inaccurate in claiming that the "individual operating under the regime of the morality of customs" is concerned with "*his* welfare" (Reginster 2018, 5; emphasis added) for—and this will be of crucial importance in what follows—under the morality of custom, one was concerned not merely with one's own well-being but at the same time with the well-being of the community as a whole (D 9). Third, as I shall show, it is simply not the case that the sovereign individual as such, in contrast to the person who is governed by the morality of custom, possesses a *moral* conscience and can thus experience *moral* guilt. What, then, distinguishes the sovereign individual with respect to the ability to experience guilt? To answer this question, I wish now to examine closely several passages from

³⁰ Reginster's full picture of the sovereign individual is richer than I let on and is by far, in my view, one of the most complex accounts to be found in the literature. He thinks that in the sovereign individual, a number of developments congeal: the bad conscience, responsibility, the value of one's worth as a person being at stake in one's moral commitments, and the "pushing back" of the sense of indebtedness into one's bad conscience, which results in the ability to feel moral guilt. Though Reginster's account is suggestive, I fear that on many occasions, it takes leave of the text.

Daybreak and *The Gay Science*. This will also throw new light on the identity of the sovereign individual.

b. Tracing the Sovereign Individual Back to *Daybreak* and *The Gay Science*

Although Nietzsche only uses the exact phrase “sovereign individual” once in his published corpus (GM II:2), prefigurations of the same idea are to be found in other works.³¹ I will start by exploring D 9, a section to which Nietzsche himself refers the reader in his discussion of the morality of custom and the sovereign individual in GM II:2. After providing us with a characterization of the morality of custom, Nietzsche goes on to say:

Those moralists, on the other hand, who, following in the footsteps of Socrates, offer the *individual* [*Individuum*] a morality of self-control and temperance as a means to his own *advantage*, as his personal key to happiness, *are the exceptions* . . . they all take a new path under the highest disapprobation of all advocates of morality of custom—they cut themselves off from the community, as unethical men [*Unsittliche*], and are in the profoundest sense evil. (D 9; emphasis in original)

And a little earlier in that section, Nietzsche writes:

The free human being [*Der freie Mensch*] is unethical [*unsittlich*] because in all things he wants to depend on himself and not on tradition: in all the original conditions of mankind, “evil” signifies

³¹ Rukgaber identifies several instances of “sovereign” and its cognates, especially in the *Nachlass*, but these are not in my view strongly relevant to our understanding of the figure of the sovereign individual in the *Genealogy* itself (Rukgaber 2012, 214–217).

the same as “individual” [*individuell*], “free,” “capricious,” “unusual,” “unforeseen,” “incalculable.” (D 9)

I think it is easy to see that in D 9, Nietzsche is concerned with a very similar figure to that of the sovereign individual: in both cases, GM II:2 and D 9, what is at issue is an individual—an *Individuum*—who, being free, is liberated from the morality of custom. Further, in both cases, importantly, a central element is that of self-control: as we saw, the responsibility that characterizes the sovereign individual requires self-control, and here again, self-control is presented as that which singles out the unethical, free individual.

But what is most remarkable about the passages quoted is that the unethical, free individual is identified here with the *Socratic* individual, the individual who is modeled after the figure of Socrates. In light of this claim, I wish to suggest that Socrates, as well as the representatives of the various schools of philosophy he engendered (“The Socratic individuals”) are for Nietzsche *paradigmatic cases of sovereign individuality*, instances of persons who overcome the morality of custom prevalent at their time and concern themselves with the individual as the locus of ethics, focusing on their own personal advantage and happiness.³²

How can we, however, make sense of the idea that the Socratic individual is a sovereign individual? Put differently, what has

³² Interestingly, in his book on the history of conscience, van Creveld notes how in Stoicism, we can find the idea of the conscience as concerned not with the gods or punishment but with the good life of tranquility that arises out of self-control over the passions. Van Creveld adds that during this time, “the position of the individual in relation to society became stronger because the community to which he belonged lost its sovereignty and, with it, some of the power it had previously exercised over each citizen separately. However, it also became weaker owing to the rise of despotic government in which he had no say whatsoever. The combination of individualism and helplessness explains the emphasis first on *syneidesis* and later on *conscientia*” (van Creveld 2015, 44). This is basically the kind of development Nietzsche traces in D 9, where the individual becomes free from the power of the community, which leads to the formation of a conscience in the individual sense. Further, the collapse of the Greek polis on the background of which Socrates emerged corresponds to the loss of communal sovereignty that van Creveld mentions here.

Socratism to do with the capacities required to produce the sovereign individual, those just distinguished? It is not my aim here to provide a complete philosophical portrait of Socrates/Socratism or defend Nietzsche's conception of Socrates in great detail, so I will limit myself to focusing on self-control and "calculative" thinking as capacities that *generate* the sovereign individual and that *further enable* him to cut himself off from the morality of custom.

I will start with the capacity for self-control, which I have already singled out as central for responsibility. Self-control, importantly, already characterizes the person who sacrifices himself for the sake of customs (D 9), but it is with the Socratic individual—the sovereign individual on my reading—that it reaches its apex. That self-control is associated by Nietzsche with Socrates and Socratism should not be surprising (e.g., D 26, 165; GS 36, 372).³³ Nevertheless, to support this claim, let me quote from A. A. Long's essay on Hellenistic ethics:

Our religious and ethical traditions have so familiarized us with the notion of self-control that it seems hard to imagine a society which had such a notion in only a tenuous form. Yet I believe this was the case with Greece prior to Socrates. It goes without saying, of course, that Greeks from Homer to Aeschylus could express and practice self-restraint. . . . But self-control, as I am using that expression, is something different. What I have in mind is a pre-Freudian notion of a self that is completely transparent to reflection, and over which its owner claims such complete authority that he finds himself in total charge of where his life is going and indulges his emotions and appetites only to the extent that he himself determines. A self with this degree of authoritativeness, transparency, and concentration is largely alien, I think, to Greek experience and conceptualization before Socrates. . . . Callicles in

³³ Notice how it is also "prudence [*Klugheit*]" that is associated with Socrates (D 26)—the trait that the social straightjacket helps produce by means of punishment, as we saw.

the *Gorgias* (491d 9) says he does not understand Socrates when he talks of “ruling oneself.” Socrates, almost certainly with irony, says he means nothing complicated but . . . “ruling the pleasures and passions within himself.” . . . As characterised by Xenophon, Socrates was “the most *enkrates* of all men over sex and bodily appetite . . .” (*Memorabilia* 1.2.1). . . . The seminal contribution of Socrates is evident not only from his decisive influence on Plato and Xenophon, but also from dominant characteristics of the other Socratic schools. Self-mastery is the keynote of the philosophies of Antisthenes, Diogenes, and even the hedonist Aristippus. (Long 1993)

We should not be alarmed by Long’s mentioning of transparency here, and I definitely do not wish to attribute to Nietzsche the view that such full transparency is possible or was actual in the specific case of Socrates and his followers; indeed, Nietzsche rejects this outright (e.g., D 116). Rather, the claim—and this is what I think Long has in mind—is that the Socratic individual *experiences* himself as possessing such self-transparency. Indeed, arguably, it is precisely *because* he is a master of self-control and has power over his “nature” (G II:2), “even in his own image,” that he falls prey to the *illusion* that his psyche is transparent to him. I will return later to another important element of the Socratic, sovereign individual which the *Gorgias* emphasizes, but now I turn to a different aspect of Socratism which, according to Nietzsche, can generate, and is characteristic of, the sovereign individual.

As we saw, the sovereign individual is adept at causal reasoning and means-ends calculations (GM II:1), but in D 544, Nietzsche writes that while

the ancient way . . . was thinking under the spell of custom [*Sittlichkeit*], for which there was nothing but established judgments, established causes, and no other reasons than those

of authority . . . it was Socrates who discovered the antithetical magic, that of cause and effect, of ground and consequence.

Socratism, embodied primarily in the figure of Socrates, with its discovery of the “magic” of causal thinking and logic, has brought about, according to Nietzsche, a diminution in the scope and force of the morality of custom, given that “in the same measure as the sense for causality increases, the extent of the domain of morality [*Sittlichkeit*] decreases [and] one has destroyed a countless number of imaginary causalities hitherto believed in as the foundations of customs [*Sitten*]” (D 10). Specifically, it is the superstitious and mythical thinking that governs the morality of custom that is undermined by Socratic thought. Precisely because he employed logic and rational thinking and did not blindly follow the herd and its gods (WS 72) but listened to the *daimonion* within—his own conscience—Socrates symbolizes for Nietzsche the “immoral” sovereign individual.

Thus, the “social straightjacket” fosters and strengthens self-control and calculative thinking, and when these processes develop far enough, they help produce the individual who is master at self-control and, through his now sufficiently powerful calculative thinking (of cause and effect, ground and consequent), breaks the authority on him of the morality of custom. Such individuals “cut themselves off from the community, as immoral men” (D 9). Notice that according to D 9, the “evil” of such individuals lies not in that they adopt or create new values.³⁴ I thus agree with Reginster that the sovereign individual is not a kind of moral “nonconformist.” And indeed, Socrates was not, despite the outrage and aggravation that he incited, a nihilist or an anarchist, and “has engaged in few activities that actually violate the traditional code” (McPherran

³⁴ This is not to deny that Socrates introduced a new “measure of value” (GM II:2), namely, that of reason: one should adhere to the community’s values because this is what is rational to do, insofar as such adherence contributes to one’s well-being. In other words, reason for Socrates is a kind of new *meta*-value.

2013, 268). Rather, the “evil” of the sovereign individual consists in that, *first*, even though he for the most part follows the values of the community, it is not tradition as such that is authoritative for him but his own conscience (governed by reason, in Socrates’s case), which is why on occasion he behaves in a manner which is perceived as “unusual” or “incalculable” (in the manner in which Socrates occasionally behaved),³⁵ and, *second*, that his motive for following the community’s values concerns his “personal advantage” (D 9) rather than the advantages of the community as a whole. These two elements are interconnected. Let me start with the second one.

With respect to personal advantage, we should remind ourselves of Socrates’s claims regarding justice and the well-being of the individual’s soul (cf. BGE 190). As Socrates makes Polus admit in *Gorgias* (477b), besides the evils that can beset one’s physical condition—“weakness, disease, ugliness, and the like”—“there’s also some corrupt condition of the soul,” namely, that of “injustice,” which is the “greatest evil there is.” Acting immorally is harmful; to be an immoral person is to suffer from a form of mental disease, a sickness of the soul. And here, according to *Gorgias*’s Socrates, lies the reason for why one should be moral: it is a form of mental health.³⁶ But I wish to suggest that it is precisely by virtue of being

³⁵ Think of Socrates remaining motionless for a long time, lost in thought, at the beginning of Plato’s *Symposium*. It is important to emphasize that the sovereign individual is *perceived* by the members of the community to be “incalculable” or “unforeseeable” or even “evil” (D 9), because he has broken free from the motivational hold of tradition and has become an individual. Besides making people *think* that he is indeed unforeseeable, this can indeed occasionally give rise to various unforeseeable actions, but it does not in any way contradict the sovereign individual’s possession of self-control, his general adherence to accepted values, and his capacity to keep his promises. And it is precisely as such an individual that he truly possesses the *right* to make promises—the pre-individualized member of the morality of custom, while “calculable” to some extent, does not possess this right fully because he is too much of a “cog” in the custom machine and cannot be answerable to himself, in contrast to the sovereign individual (GM II:3). The sovereign individual *is* responsible for himself, and when he gives his word, this has weight. I thank an anonymous reader for pushing me to further clarify this point.

³⁶ Thus, Gregory Vlastos: “the question ‘Why should I be moral?’ . . . is for all Greek moralists a perfectly proper and unavoidable one. . . . They agree that the right reply is ‘Because moral conduct offers me the best prospect for happiness’” (Vlastos 1991,

concerned with her own personal happiness or *eudaemonia* that the sovereign individual, on this view, could be seen as possessing her own personal conscience, one that is separate from the community's. To see this point more clearly, I wish to return to D 9, where Nietzsche writes:

Everywhere that a community, and consequently a morality of customs exists, the idea also predominates that punishment for breaches of custom will fall before all on the community: that supernatural punishment whose forms of expression and limitations are so hard to comprehend and are explored with so much superstitious fear. The community can compel the individual [*Einzelnen*] to compensate another individual or the community for the immediate injury his action has brought in its train; it can also take a kind of revenge on the individual for having, as a supposed after-effect of his action, caused the clouds and storms of divine anger to have gathered over the community—but it feels the individual's guilt [*Schuld*] above all as *its own* guilt and bears the punishment as *its own* punishment.

What characterizes the morality of custom with respect to the feeling of guilt, then, is that in the age of the morality of custom, one's guilt was *not* absolutely and solely one's own; rather, it was seen as belonging to the community as a whole: "‘customs have grown lax,’ each wails in his soul, ‘if such actions as this are possible’" (D 9). In other words, at this early stage, there is no sense of *personal responsibility*; there is no sharp line dividing the individual's conscience from the community's—the individual cannot thus experience his guilt as exclusively his own. He thus cannot, at this early stage, "stand security for [himself]" (GM II:3)—be fully answerable

203–204). Vlastos goes on to defend the view that for Socrates, virtue is necessary and sufficient for happiness, but other goods, when governed by virtue, could also contribute to one's happiness.

to *himself*. A similar set of ideas is presented in GS 117, a section that is usually ignored in discussions of the sovereign individual despite its immediate relevance:

Herd remorse [*Heerden-Gewissensbiss*].—During the longest and most remote periods of the human past, the sting of conscience [*Gewissensbiss*] was not at all what it is now. Today one feels responsible only for one's will and actions [*Heute fühlt man sich nur verantwortlich für Das, was man will und thut*], and one finds pride in oneself. . . . But during the longest period of the human past nothing was more terrible than to feel that one stood by oneself. To be alone, to experience things by oneself, neither to obey nor to rule, to be an individual [*Individuum*].—that was not a pleasure but a punishment: one was sentenced "to individuality" [*zum Individuum*"]. Freedom of thought was considered discomfort itself. . . . To be a self and to esteem oneself according to one's own weight and measure—that offended taste in those days. . . . In those days, "free will" [*freie Wille*"] was very closely associated with a bad conscience [*das böse Gewissen*]; and the more unfree one's actions were and the more the herd instinct rather than any personal sense found expression in an action, the more moral [*moralischer*] one felt. Whatever harmed the herd, whether the individual [*Einzelne*] had wanted it or not wanted it, prompted the sting of conscience in the individual [*dem Einzelnen Gewissensbisse*].—and in his neighbor, too, and even in the whole herd.—There is no point on which we have learned to think and feel more differently.³⁷

³⁷ Admittedly, it is not so clear what specific historical example of a community Nietzsche has in mind where there is a collective sense of guilt or bad conscience but not an individual one. Ancient Thebes, per *Oedipus Rex*, is suffering from a plague—a collective divine punishment for the murder of King Laius. There is no mention in the play, however, of collective guilt for the crime. If we are to rely on Williams (1993b), the ancient Greeks—and Oedipus himself, the murderer of Laius, is perhaps an example—were prone to feelings of shame rather than guilt.

First, to support my claim that here, too, Nietzsche is concerned with the sovereign individual, let us notice the similarity of some of the ideas in this section to those in D 9 and to the characterization of the sovereign individual in GM II:2: in all three passages, Nietzsche is concerned with an *Individuum* who breaks free from the morality of the community. Further, while here, in GS 117, individuality is explained in terms of possessing “one’s own weight and measure [*sich selber nach eigenem Maass und Gewicht schätzen*],” in GM II:2, the sovereign individual is characterized as someone who “possesses his *measure of value* [*sein Werthmaass*].” Third, in both GS 117 and GM II:2, the individual is described as possessing a sense of *pride* in himself.

Finally, notice that just as in D 9, Nietzsche in GS 117 emphasizes a difference with respect to the experience of guilt or pangs of conscience: under the morality of custom, “whatever harmed the herd” and went against customs gave rise to feelings of guilt *across* the community. In contrast to today, where one feels “responsible only for one’s will and actions,” during the “most remote periods of the human past,” where the morality of custom ruled, one did not feel responsible only for one’s own actions but felt responsible for the actions of every member of the community. The resulting picture is that one felt responsible for the actions of others, and others also felt responsible for one’s own actions. In other words, *under the governance of the morality of custom, responsibility is not individualized*. There is thus not yet a sense that one is really *individually* responsible for one’s actions; there is rather a sense of shared responsibility for every infraction committed. The implication of this is that one’s concern is not merely with one’s own well-being and “personal advantage” (D 9), but one sees one’s own well-being as tied up with the well-being of the whole. This, of course, is perfectly coherent with the basic idea of the morality of custom, where a tradition is “above all directed at the preservation of a community, a people” (HH 96).

The conclusion to draw is that one central feature of the sovereign individual is precisely that he possesses his *own individual*

conscience. This means, first, that he is not prone as a matter of general rule to feel guilt for the misdeeds of others in his community; he feels guilt—in this sense—only for his own actions since he takes himself to be responsible for *his* actions alone. Second, when responsible, the sovereign individual bears the *entire* burden of responsibility for his actions: he does not conceive of his fellow community members or anyone else as sharing in his moral load as per the morality of custom. This—the *personal, individuated* sense of responsibility—is the extraordinary responsibility that the sovereign individual experiences and, in virtue of his self-control, feels entitled to. As Nietzsche puts it: “What will he call this dominating instinct [of extraordinary privilege of responsibility], supposing he feels the need to give it a name? The answer is beyond doubt: this sovereign man calls it his *conscience* [*sein Gewissen*]” (GM II:2). And though Nietzsche emphasizes the word “conscience” here, I think that the possessive pronoun he uses here as well as right at the beginning of the next section (“His conscience? [*Sein Gewissen?*]”; GM II:3) is telling. The members of the primordial herd, given that they were prone to experience pangs of conscience, possessed a conscience, too. But the sovereign individual has his *own*, individualized conscience, a sovereign conscience, namely, one that is “set apart” and “independent” (GM III:5).³⁸ Possessing his own sense of personal responsibility, the sovereign individual is thus subject to his own, individual “feeling of guilt, of *personal* obligation” (GM II:8; emphasis added). It is only once the individual considers himself responsible for his own actions that he can experience guilt in a sense familiar to us nowadays: the torture of self-accusation which involves the belief that one’s misdeeds are one’s *own* individual fault (rather than the fault of the community or the

³⁸ It is interesting to note that while Nietzsche seems to think that “Today one feels responsible only for one’s will and actions” (GS 117), this is not exactly accurate, since the possibility of collective guilt is something that is still open to us. Nevertheless, Nietzsche thinks that the sense of personal responsibility presupposes and developed out of collective guilt.

gods). Finally, and third, the sovereign individual regards his own conscience as authoritative—not the morality of custom, anchored as it is in reverence for the tradition or the gods. It is in all these different ways that the sovereign individual becomes “liberated” from the morality of custom.

It is illuminating to focus on an important distinction Nietzsche consistently makes both in GS 117 and in D 9 in the numerous passages I cited above when discussing the morality of custom: a distinction between the single person (*Einzelne*) and the individual (*Individuum*): whereas the former concept refers to a single member of the community who has not yet broken away from the morality of custom and feels guilty precisely *for* such attempts at individuality (a guilt he shares with other members of the community), the latter concept refers to the sovereign individual who no longer feels guilty about his own sense of personal responsibility but rather feels proud about possessing it. Nevertheless, the sovereign individual continues to feel guilty for his own personal misdeeds, for which he considers himself solely responsible. The former obeys *customs* out of reverence: respect, on the one hand, and fear regarding the consequences of unethical action for the *entirety* of the community, on the other. Failure to obey results in *shared guilt*: self-laceration for disobeying tradition/the gods that involves an element of fear of impending catastrophe. The latter, in contrast, obeys one’s *own conscience* out of pride and concern for *one’s own well-being*. Failure to obey results in *personal guilt*: self-laceration for failing to do what one considers to be prudential and which involves, in addition, an element of fear for one’s well-being.

But what do the self-control and the calculating abilities mastered by the sovereign individual have to do with what I have called individualized conscience? First, the more an agent possesses self-control and is aware of such self-control, the more, that is, he comes to possess a “proud consciousness . . . of his power . . . over himself” (GM II:2), the less the agent will tend to ascribe the performance of his actions to other agents. The suggestion is that for

us to feel *individual* guilt is to consider ourselves as *individually* responsible for our deeds, to think that the misdeed performed was up to us and only up to us, that *we* were in control of ourselves and our actions, not the community or the gods. Second, the more an agent's capacity to think critically and grasp causal connections grows, the less he will be attracted to the view that there are invisible forces at work that determine one's actions from the outside (the community, the gods). These two capacities, it is easy to see, support each other to produce an individual who can experience guilt in the individualized sense; an individual who recognizes the deed as his own and thus takes the entire burden of guilt on himself.

These comments are, of course, related to how we should understand Nietzsche's talk of the sovereign individual's autonomy in this context: "autonomous" (GM II:2) need not mean something unfamiliar,³⁹ and does not require freedom as a *causa sui*, but means the ability to control oneself in accordance with what is expected and independently of immediate external sanctions and guidance. The sovereign individual has internalized the values of his society and is able to control himself and act reliably in light of these values. But this is what *responsibility* is all about.⁴⁰ And yet this responsibility is incomplete as long as it is not *individualized*. With the sovereign individual, however, it is fully consummated.

A concrete manifestation of the sovereign individual's individualized responsibility is to be found in Nietzsche's claim that the sovereign individual knows himself to be "strong enough

³⁹ Leiter claims that "Nietzsche does not think human beings have a capacity for genuinely autonomous choice, so, as is often the case, Nietzsche is [in his description of the sovereign individual] using familiar words in unfamiliar senses" (Leiter 2002, 227–228). What the unfamiliar sense is Leiter does not fully explain, but he adds that at the end of the day, "autonomy" in this context merely comes down to having the right to promise (228).

⁴⁰ A closely related account is given by R. J. Wallace, where a morally responsible agent is one who possesses "reflective self-control." Reflective self-control involves capacities of the following two kinds: "(1) the power to grasp and apply moral reasons, and (2) the power to control or regulate [one's] behavior by the light of such reasons" (Wallace 1994, 157).

to maintain [his commitments] even “in the face of fate” [*gegen das Schicksal*]” and possesses “power over himself and over fate [*das Geschick*]” (GM II:2). I wish to suggest that the concept of fate Nietzsche has in mind here is most plausibly the ancient Greek notion of fate to be found in Homer and in Greek tragedies where fate is understood as a force—occasionally identified with Zeus’s will (see Duffy 1947)—that governs one’s actions in a way that lies beyond one’s control. This is, however, precisely how an individual might feel who is *not* sovereign and thus does not know how to plan ahead or to control himself and still believes in the superstitions of his community; he might explain his actions, especially his *misdeeds*, as resulting from some form of fateful divine intervention (“‘He must have been deluded by a *god*’”; GM II:23).⁴¹ Presumably impressed with Plato’s view according to which humans are playthings of the gods,⁴² Nietzsche held that “for the longest time these Greeks used their gods precisely so as to ward off the ‘bad conscience’”; that is, they employed “the conception of gods” to attribute responsibility for their misdeeds to the gods and thus shuffle off at least some of their own responsibility. As Nietzsche says, “in those days [the gods] took upon themselves . . . the guilt” (GM II:23). This presumably has the effect of minimizing not only one’s sense of personal responsibility but the intensity of one’s self-torture as well; rather than direct its entire ferocity at himself, the subject believes it is not really all his fault but a

⁴¹ Thus, *King Lear*’s Edmund: “This is the excellent foppery of the world, that when we are sick in fortune, often the surfeits of our own behavior, we make guilty of our disasters the sun, the moon, and stars; as if we were villains on necessity; fools by heavenly compulsion” (Act I, Scene 2).

⁴² Plato: “man, as we said before, is only a plaything constructed by the gods” (*Laws*, 803b–d); and “let us suppose that each of us living creatures is a puppet constructed by the gods, we do not know whether it is a plaything for them or to some serious end, but we do know that these sensations inside us are like sinews or cords that drag us along and, being opposed to each other, pull us to opposite actions; and here it is that lies the line dividing virtue and vice” (*Laws*, 644d). Nietzsche was thoroughly familiar with Plato’s *Laws* as well as with these specific passages (see D’Iorio 2016, 105–106).

god's.⁴³ Interestingly, alongside this exculpating use of the concept of the gods, one can find an echo of that other, earlier period, where the central self-criticism for the misdeed was one of prudence, for here, too, according to Nietzsche, the Greek gods, who are after all a human invention, complain that mortals “*make themselves wretched through folly*” (GM II:23, Nietzsche quoting Homer; see also D 78). In other words, the ancient Greeks were ambiguous between accepting at least some responsibility for their misdeeds—attributing it to their own folly—and shifting responsibility away and onto the gods by *explaining* their folly as itself originating in divine mischief. From the human perspective, the gods are to blame; from the gods’ perspective, which is in essence another human perspective, humans act foolishly and thus are to blame at least to some extent.⁴⁴

The sovereign individual, in contrast to those Greeks who used the gods to shirk their responsibility, does not regard himself as a plaything of the gods or as overpowered by fate but considers himself as possessing the “extraordinary privilege of *responsibility*” (GM II:2). So in his discussion of fate, Nietzsche does not mean to attribute to the sovereign individual some libertarian free will (or some such conception of free will); the sovereign individual has the capacity for self-control and is released from the view that the *gods* or any other agents are in control of his actions. To that extent, the sovereign individual acquires power over “fate.”⁴⁵

⁴³ Agamemnon is an example that Williams discusses (Williams 1993b, 52). See also HH 45 for an earlier formulation of the same idea.

⁴⁴ But in a glaring contradiction to this account, we have the example of Oedipus, who gouged out his own eyes upon realizing his crime; no guilt, it seems, was shuffled off to the gods in this case. In his discussion of Greek guilt in *Daybreak*, Nietzsche claims that “antique sensibility commented” upon the tragic hero’s misdeed in the following way: “Yes he should have gone his way a little more cautiously and with less haughtiness!” (D 78). But this is definitely far too mild a response compared with the one familiar from Greek tragedies.

⁴⁵ I thus disagree with Leiter when he implicitly identifies fate with determinism and claims that given that Nietzsche in general holds that we lack libertarian free will, the sovereign individual is “delusional” in thinking that he has power over fate “if he *really* believes any of this” (Leiter 2011, 109). But from the sovereign individual’s perspective,

To my analysis, it might be objected that the figure of the sovereign individual as construed here makes a bad fit for the figure of Socrates, for Socrates regularly deferred to the authority of myth and the gods and thus could not be said genuinely to possess his own conscience. But this would be to misconstrue Socrates's characteristic critical attitude toward the deities, where Socrates typically "takes it to be obligatory to subject extra-rational signs to rational interpretation and confirmation whenever possible" (McPherran 2013, 267). As Nietzsche puts it, Socrates's task was that of "*putting the god to the test in a hundred ways to see whether he has told the truth*" (WS 72). Further, while Socrates does not explicitly proclaim or theorize about the possession of a personal conscience, his famous *daimonion*, the "internal, private admonitory 'sign'" and "voice" (McPherran 2013, 268) that makes its appearance (among other places) in the *Apology* and *Phaedrus* and warns him not to pursue certain disadvantageous courses of action, bears a striking resemblance to what we understand by "the voice of conscience."⁴⁶

To conclude this discussion, I first return to the question regarding the role of the sovereign individual in Nietzsche's genealogy of bad conscience and guilt. In my view, extant accounts of the sovereign individual fail, on the one hand, to capture the novelty that the sovereign individual introduces and, on the other hand, overinflate the sovereign individual's accomplishment. Let me explain the latter point.

it is rather he who believes that forces of fate (in the Greek sense) govern his life who is delusional!

⁴⁶ A central difference between the Socratic *daimonion* and the phenomenon of the feelings of guilt is that in the Platonic texts, there is no occurrence where the *daimonion* reproaches subsequent to the performance of the action; rather, it functions purely *prospectively*. To the extent that Nietzsche is implicitly suggesting that Socrates and his followers possessed an individual *retrospective* admonishing conscience as well, he goes beyond what the Platonic texts themselves license. There is, however, an exception of sorts to the purely prospective function of the *daimonion* where Socrates at his trial claims "that he was 'aware of himself' that he was innocent: that is, that his conscience was clear" (van Creveld 2015, 32). If this is true, then his conscience not only warned him but also reassured him that what he did was right.

Zamosc attributes to the sovereign individual the capacity to “feel morally responsible” (Zamosc 2012, 123), and Reginster wishes to claim that with the ideal of the sovereign individual, we have a mental structure that makes *moral guilt* possible. To be a sovereign individual, for Reginster, involves caring about one’s own self-worth where failure to meet the commitments one takes upon oneself can result in moral guilt: a diminution in one’s self-estimation corrected by self-laceration (Reginster 2018, 16–18). If my account of the sovereign individual as a Socratic individual is on the right track, however, then the sovereign individual is motivated by a mixture of pride (or respect for himself) and concern for his own well-being (happiness or the well-being of his soul). But this goes against Reginster’s claim that the sovereign individual is *not* concerned with “his welfare” (2018, 5), and it means, moreover, that the sovereign individual is still motivated by prudential considerations that do not amount to genuine moral concern. It is thus neither the case that the sovereign individual “cares about keeping his promises ‘as such’” (Reginster 2018, 6, quoting Nietzsche, GM II:14) nor that he is a genuine “*morally liable author*” of his actions (Zamosc 2012, 123).⁴⁷

But if it is the case that the sovereign individual is not yet able to experience genuine moral guilt, then it is not clear where in the *Genealogy* is to be found an explanation for this phenomenon, the phenomenon, namely, where an agent lacerates himself or herself precisely *for doing something morally wrong*, that is, for finding his or her action wrong *as such* and not because the gods proclaimed it wrong and one fears their wrath or because immorality corrupts one’s soul. For a full explanation of the emergence

⁴⁷ Reginster’s claim that the sovereign individual is concerned with keeping his promises “as such” is perplexing on Reginster’s own understanding of the sovereign’s psychology as well, for, on his picture, the sovereign is at the end of the day concerned with his own worth as a person, namely, as someone who can live up to the commitments he has taken upon himself, that is, with his power over himself, *not* with promise keeping “as such.” For this individual, it seems, promise keeping has no value in and of itself; it is merely an occasion to test one’s power over oneself.

of the phenomenon of moral guilt, we will have to go beyond the *Genealogy* for answers.

Where does this leave us, however, with respect to the question of Nietzsche's evaluation of the sovereign individual? Is this an ideal Nietzsche erects or a figure he wishes to mock and knock down? This debate, in my view, often takes the form of a false dichotomy, where the sovereign individual is seen as representing either a "target" or an "ideal" (cf. Rukgaber 2012, 214). We can now see why this figure has generated such conflicting evaluations on the part of scholars, for the ambiguity in Nietzsche's evaluation of the historical figure of Socrates carries over to his ambiguous evaluation of the sovereign individual. On the one hand, Nietzsche claims in GS 340 that he "admires the courage and wisdom of Socrates in everything he did and said—and did not say," but on the other hand, already at the time of *The Birth of Tragedy* and all the way to the end in "The Case of Socrates," he perceives his excessive rationality as a symptom and cause of decadence. On the one hand, Nietzsche admires his individuality, his standing apart, his free-spiritedness (HH I:433); but on the other, he vilifies him for introducing dialectics and thus defeating the "noble taste" (TI, Socrates, 5) of the Greeks. Moreover, Nietzsche raises the suspicion that Socrates was motivated by *ressentiment* and wanted to take revenge "on the nobles he fascinated" (TI, Socrates, 6). It thus seems that it is precisely Socrates's fine intellect (WS 86), precisely those capacities for logical and causal thinking that give rise to sovereignty and are a paradigmatic expression of sovereignty, that, at the same time, signify for Nietzsche in the figure of Socrates a decline and a danger: "rationality at any cost, a cold, bright, cautious, conscious life without instinct, opposed to instinct, was itself just a sickness" (TI, Socrates, 11).

Indeed, "The Case of Socrates" is so ruthlessly negative that given the rather positive manner in which Nietzsche describes the sovereign individual in the *Genealogy*, "The Case of Socrates" might be taken as strong reason to think either that Socrates was

never an instance of sovereignty for Nietzsche or that by the time of *Twilight of the Idols*, Nietzsche had revalued sovereignty, or at least Socrates's embodiment of sovereignty, and regarded it as thoroughly negative. These conclusions would be wrong. It is important to keep in mind that even here, in *Twilight*, Socrates is presented as someone who, despite the "anarchy of his instincts" (TI, Socrates, 4), has "mastered them all" (TI, Socrates, 9). Thus, Socrates still presents us with a paradigmatic case of sovereignty: it speaks rather to the supreme power he had over himself that he could master himself precisely at this most difficult of times, when Greek society was disintegrating all around him. Further evidence for the view that the Socratic individual is a prime case of sovereignty—and this highlights again how ambiguous the phenomenon of the sovereign individual is for Nietzsche—is to be found in a section from *The Gay Science*. Here Nietzsche discusses the "signs of corruption"—precisely the collapse of the morality of custom at the background of his discussion of Socrates in *Twilight* ("Athens was coming to an end," TI, Socrates, 9)—and claims that when "customs decay" (*die Sitten verfallen*; translation altered),

those men emerge whom one calls tyrants: they are the precursors and as it were the precocious *firstling instances of individuals* [*Erstlinge der Individuen*]. Only a little while later this fruit of fruits hangs yellow and mellow from the tree. (GS 23; Nauckhoff's translation)

We should recall how Nietzsche uses the same image of the ripest fruit to describe the sovereign individual in GM II:2. We should also recall that Socrates is described in "The Case of Socrates" as a tyrant: "dialectics," of which Socrates was a master, "lets you act like a tyrant" (TI, Socrates, 7). Finally, in an unpublished note from 1888, where Nietzsche discusses the fateful consequences of attempting to tear moral judgments out of their "Greek and Greek-political ground and soil" and base them on "dialectics"

and “reasons,” he claims that the result is the “perfectly absurd ‘individuum’ [*Individuum*] in itself!” This individuum, he clarifies, is a “degenerate type of man—‘the good man,’ ‘the happy man,’ ‘the wise man’—*Socrates* represents a moment of the profoundest perversity in the history of values” (KSA 1888, 14 [111]; WP 430; emphasis added).⁴⁸

VI. Beyond the Sovereign Individual

I will now go beyond Nietzsche’s second essay and connect his account of guilt and bad conscience with some ideas raised in the *Genealogy*’s first essay. This will take us from sovereign responsibility to an even deeper sense of responsibility and, accordingly, to a deeper conception of guilt. The sovereign individual, as reconstructed above, considers himself individually responsible for his deeds, and when he fails to act in accordance with the values he’s internalized, he feels individualized guilt. How, however, does he conceive of his failure? Since the gods and the community are off the hook, the performance of the misdeed cannot be ascribed to them. But if he is the only one to blame, how is this blame to be thought of? The following step in the genealogy provides an answer to this question, for at the end of this next development, the agent not only feels the pangs of conscience after having performed a misdeed but also thinks, “*I am to blame, and should feel guilty, for I could have acted differently.*” More precisely, the thought of leeway freedom now serves to *justify* the agent’s own guilt from the agent’s own perspective (as well as moral condemnation and ascriptions

⁴⁸ Further evidence for the ambiguity of the figure of the sovereign individual and Socrates/Socratism is to be found in BGE 262, which again picks up the theme of corruption and claims that the “individual [*Individuum*] . . . obliged to give himself laws and to develop his own art and wiles for self-preservation” is a symptom of the collapse of the community. That Nietzsche has Socrates in mind in this section as well as a prime example of the *Individuum* can be plausibly surmised on the basis of his mention of the Greek polis at the very beginning of the section.

of moral guilt by others). And indeed, for Nietzsche, the idea of leeway freedom as a justification for punishment is a rather late psychological development: the now “natural” and “unavoidable idea” that “the criminal deserves [*verdient*] punishment *because* he could have acted differently’—is in fact an extremely late and subtle form of human judgment and inference” (GM II:4).

The transition to this deeper sense of responsibility can be analyzed into two dimensions, one subjective-psychological and the other conceptual. First, as we know, the sovereign individual possesses self-control and an awareness of his power over himself. But, as Rée explains, this belief in our capacities generates the illusion that because we possessed at the time of action the *general* ability to do the right thing, we also could have *exercised* these capacities at the time of action, at the time of our misdeed. The sovereign individual, subsequent to this development, will therefore be in a position to say to himself, “At that time in the past, not only (1) did I possess the power to do the right thing, but (2) I also could have employed that power and acted in the right way. Therefore, the entire burden of responsibility, and consequently the guilt, is mine.”

We see here how Nietzsche’s and Rée’s accounts can complement each other. Rée supplies us with an explanation of how the notion of leeway freedom arises in us once we conceive of ourselves as possessing certain agential capacities. On the other hand, Nietzsche provides us with a genealogy—missing in Rée—of how these capacities are acquired in the first place.

To further reinforce this subjective illusion, we have to make the second, conceptual step. This involves the idea of metaphysical or libertarian free will. This notion is needed to provide conceptual support to the thought of leeway freedom, for, arguably, without it, the idea that despite the conditions that held at the time of action one could have nevertheless exercised one’s general capacities and done the right thing would make less sense upon reflection. Nietzsche explains the genesis of this idea in the context of his analysis of the *ressentiment*-driven slave revolt in morality in the

first essay of the *Genealogy*. There he also draws an important connection between the notion of free will and the concept of the self or the neutral *subject*, free to do one thing or another at any given practical juncture. Here is the relevant important passage (I quote him at some length):

To demand of strength that it should *not* express itself as strength, that it should *not* be a desire to overcome, a desire to throw down, a desire to become master, a thirst for enemies and resistances and triumphs, is just as absurd as to demand of a weakness that it should express itself as strength. A quantum of force is equivalent to a quantum of drive, will, effect—more, it is nothing other than precisely this very driving, willing, effecting, and only owing to the seduction of language (and of the fundamental errors of reason that are petrified in it) which conceives and misconceives all effects as conditioned by something that causes effects, by a “subject,” can it appear otherwise. For just as the popular mind separates the lightning from its flash and takes the latter for an *action*, for the operation of a subject called lightning, so popular morality [*Volks-Moral*] also separates strength from expressions of strength, as if there were a neutral substratum [*indifferentes Substrat*] behind the strong man, which was *free* to express strength or not to do so. But there is no such substratum; there is no “being” behind doing, effecting, becoming; “the doer” is merely a fiction added to the deed—the deed is everything. The popular mind in fact doubles the deed: when it sees the lightning flash it is the deed of a deed: it posits the same event first as cause and then a second time as its effect . . . no wonder if the submerged, darkly glowing emotions of vengefulness and hatred [i.e., *ressentiment*] exploit this belief for their own ends and in fact maintain no belief more ardently than the belief that *the strong man is free* to be weak and the bird of prey to be a lamb—for thus they gain the right to make the bird of prey *accountable* for being a bird of prey

[*das Recht, dem Raubvogel es zuzurechnen, Raubvogel zu sein*].
(GM I:13).

One way to read this passage is the following. In order to take revenge upon the masters, the slaves need to be able to think of the masters as accountable for their evil actions. This requires the idea that the masters *can act differently*, that is, are free to choose not to express their strength but to act more like slavish “lambs”—it requires the idea of the free subject, the subject of leeway freedom. This idea exploited the already extant (mis)conception that there is such a neutral (neither strong nor weak) self behind the action, behind the deed—an idea that people are confused into accepting as a result of the “seductions” of reason and grammar. The agential formation which is the sovereign individual thus further develops with the notion that the *free self* bears the entire responsibility for one’s actions. Subsequent to a performance of a misdeed, the subject can now torment herself for not acting differently, given that she indeed *could have* done so. If, at the beginning of the genealogy we have been examining, punishment was not meted out on the basis of any robust conception of responsibility, here, at the end of the trajectory, we have the notion that people are to be punished—by others *or* by themselves—whenever they go against the values or laws of a community because they deserve to be, for, after all, they could have acted differently (GM II:4). How, however, can the self freely choose to act this way or that without being determined to act in any given way? What is it that determines it to act if it is indeed “neutral”? Here we find again—though implicitly—the idea of the self as a *causa sui*: the self that enjoys leeway freedom and is thus responsible for its actions is not caused to act by external forces but causes itself to act in a free manner.

We still have to address the problem of the *as such*: how does a conscience that tortures the agent for doing something that is morally wrong form? Here, too, section 13 of the first essay offers

potential help. Under master morality, actions were determined “good” or “bad” dependent on *who* performed them; no action was considered bad as such. In contrast, with the emergence of slave morality—with morality as we now know it—the idea emerged that certain kinds of actions are wrong regardless of who performs them. But this development, Aaron Ridley has argued, requires the distinction between the doer and the deed, which, as we learn from section 13, the slave revolt brings in its train. This distinction precisely helps make room for the possibility of evaluating kinds of actions independently of the kind of person who performs them: “To condemn and repudiate an action purely on the grounds that it is an action of a certain *kind* is to condemn it no matter whose action it is, and is thus to evaluate the deed in isolation from the doer” (Ridley 1998, 32).

But this is of no help with respect to our problem, for even if the doer-deed distinction is necessary for the possibility of an evaluation of an action as such, this distinction, in itself, does not explain how the latter evaluation *itself* emerges. Indeed, it is hard to imagine a community where no such distinction between doer and deed is present. As I just explained, under master morality, this distinction was operative, for only by means of it could one determine whether an action was “bad” or not; it was necessary to know both who did it (the identity of the doer) and the nature of the action (the deed) before one could come to judge the value of the action. And yet this evaluation was not of the action *as such*. Further, to consider an action as wrong regardless of the identity of the person who performed it is not to consider it as wrong as such in the relevant moral sense, for, as I explained, this also requires abstracting one’s evaluation of the action from its negative consequences.

We can appeal to another aspect of the slave revolt for assistance. The slaves in the slave revolt in morality, motivated by *ressentiment*—the desire to take revenge upon the masters—create with the help of the priests new values by means of which they can

morally condemn the masters.⁴⁹ This involves a revaluation of the old masterly ways of behaving (formerly praised as “good”) as “evil” and of the slavish traits and behavior (formerly denounced by the masters as “bad”) as “good.” In the attempt to persuade the slaves to adopt the new values, this reinterpretation appeals to God (cf. GM I:7), specifically the *Christian* God. But do the slaves, then, consider the types of actions now valued as “good” to be good as such, or do they consider them good only out of the belief that doing so will secure for the virtuous the reaping of divine rewards and the avoidance of divine wrath? In GM I:14, Nietzsche represents the slaves as believing, in light of the success of their revolt, not only that they are “better” but that they are “better off,” “or at least will be better off someday.” This indicates that the slaves come to believe *both* that the new forms of behavior praised as “good” are intrinsically good, that is, good as such, *and* that endorsing them is prudent. For instance, being humble is intrinsically good; it makes you a better person. God loves those who are humble, and he loves them because they manifest a trait that is intrinsically good.⁵⁰ Conversely, being haughty or violent is intrinsically bad; it makes you an intrinsically worse person. Moreover, this new morality, since originating from the one God, before whom all are equal, applies equally to everyone, *regardless of who one is*. As a result, in light of the slave revolt, certain courses of action are good or bad *as such* in both of the senses distinguished here. At the same time, however, the adherents of the slave morality are made to understand that the new championed ways of behaving are instrumentally good, that is, have beneficial consequences, for God will reward those who abide

⁴⁹ It is important to underline that the slaves do not adopt the new values *in order* to take revenge on the masters; the process of the creation and adoption of the new values involves a form of *self-deception*. For more on this, see Elgat 2017, chap. 4.

⁵⁰ Notice that it is not God’s loving these values that makes them good, but God’s loving them is nevertheless a reason for thinking that they are good in themselves; for God, after all, is an expert in goodness.

by his commands: the meek shall inherit the earth and enjoy the blessedness of life in heaven.

This means that we have not yet reached our goal. For one, the intrinsic valuation or disvaluation of certain courses of action is still too closely tied to the belief in God. Genuine moral guilt, we might want to say, concerns feeling bad about doing something that is considered wrong independently of any belief in God. Second, the new morality is still (at least partly) understood from the utilitarian perspective. Nietzsche indeed suspects that under Christianity, we have a subject that is motivated to adhere to morality mostly by personal, utilitarian considerations. Indeed, the Christian, whom Nietzsche gives as another example of a sovereign individual, “‘considered first of all his own salvation’” (D 9).

Let us take, then, the final step in our genealogy and turn again to *Daybreak* for answers. Here, in a section titled “The Echo of Christianity,” Nietzsche writes:

That men today feel the sympathetic, disinterested, generally useful social actions to be the *moral* [*moralische*] actions—this is perhaps the most general effect and conversion which Christianity has produced in Europe: although it was not its intention nor contained in its teaching. But it was the residuum of Christian states of mind left when the very much antithetical, strictly egoistic fundamental belief in the “one thing needful,” in the absolute importance of eternal *personal* salvation, together with the dogmas upon which it rested, gradually retreated and the subsidiary belief in “love,” in “love of one’s neighbour,” in concert with the tremendous practical effect of ecclesiastical charity, was thereby pushed into the foreground. (D 132)

The idea Nietzsche is introducing here is that once the belief in the Christian God has died (GS 343) along with all its religious-metaphysical appurtenances (life in heaven, the immortality of the soul, sin, etc.), the personal utilitarian considerations that governed

the believer's adherence to the faith have lost their grounding, and what we have been left with is the "residuum Christian states of mind" and the "subsidiary belief" in "love of one's neighbor," that is, the positive feeling for Christian *moral* value and the adherence to it themselves. Lacking their personal, instrumental, otherworldly justification, they become purely moral.⁵¹ In other words, here, too, we seem to have a very Rée-ian kind of explanation: norms or values are adopted (partly or mostly) for their utility but become over time psychologically ingrained in such a way that once the belief in their utility is shed, they nevertheless remain affectively, doxastically, and practically effective. We now finally have before us the truly ripe fruit of this long process: the sovereign individual, the product of the morality of custom, having passed through the crucible of Christianity and, with its collapse, disabused of his egoism, can now experience individual, *moral* guilt for his moral misdeeds.

VII. Christian Guilt

To round off our discussion of Nietzsche's genealogy of guilt, we have to briefly address his explanation of the emergence of *Christian* guilt. The Christian notion of guilt goes further than "mere" guilt in that it involves two central ideas: first, the idea that *all* suffering is a punishment for some wrong deed or other (GM III:20; D 78); second, the idea that a human is a sinful being and is guilty *as such*—a radical version of the idea of ontological guilt. According to this latter idea, on which I will focus, it is not merely that one's actions are "evil," but they are expressions of one's fallen nature, one's "sinfulness" and "depravity" (GM III:16; see also GM

⁵¹ Though here, too, in D 132, Nietzsche admittedly goes on to explain how the effort was then made to justify those Christian states of mind, values, and actions in terms of *social* utility, which, however, is supposed now to have moral value.

II:22). Now, one thing that is interesting in Nietzsche's account is the following. As we saw, for the philosophers who belong to the metaphysical tradition, ontological guilt provides the transcendental condition of justifiability, as well as the explanatory ground for factual or empirical guilt, and in this way comes *first* in the order of being and explanation. For Nietzsche, in contrast, this guilt comes into the world *last*. This fits Nietzsche's general view of how philosophers usually go about things. Philosophers, he says,

confuse what comes first with what comes last. They take what comes at the end . . . the "highest ideas" . . . the last wisps of smoke from the evaporating end of reality—and they put it at the beginning, *as* the beginning. But again, this is just their way of showing respect: the highest should *not* grow out of the lowest, it should *not* grow at all. . . . Moral: everything from the first rank must be a *causa sui*. (TI, Reason, 4).

Now, the Christianization of guilt is a product of what Nietzsche refers to with the expression "the moralization" of debt (*Schuld*) and duty, which he explains as "their pushing back into the conscience; more precisely, the involvement of the *bad* conscience with the concept of god" (GM II:21). Though presented somewhat confusingly, the idea here, I think, is quite straightforward. We saw that the second interpretation of the debtor-creditor relationship involved the concept of debt to the tribe's ancestors—a debt one has a duty to pay. We also saw that these ancestors, in the course of time, turn in the community's imagination into all-powerful gods. Now, Nietzsche claims that "the advent of the Christian God" represents "the maximum god attained so far" and was consequently accompanied "by the maximum feeling of indebtedness [*Schuldgefühls*] on earth" (GM II:20). This in itself—that is, the having of this debt as well as the awareness of having this debt (*Schuldbewusstseins*)—does not give rise to any guilt. As I have claimed, there is nothing wrong in merely being indebted; it

is when one *fails* to abide by one's obligations that one can acquire (empirical) guilt.

So what does Nietzsche mean by the *moralization* of this immense, indeed, infinite and thus "irredeemable" (GM II:21), debt?⁵² How does the bad conscience become involved with the concept of God? What Nietzsche has in mind here is the development whereby the bad conscience comes to have this infinite debt toward God as its *object*. "Guilt before God" (GM II:22) is guilt not for any specific misdeed but for *being* indebted in itself; the very state of indebtedness is now perceived as something wrong, something *morally* wrong (wrong as such). As a result, man perceives his very *being* as wrong. Specifically, man condemns his very being as a finite creature with "ineluctable animal instincts": he conceives of his own "absolute unworthiness" (GM II:22) or "worthlessness" (Diethe's translation), conceives of himself as wretched, vile, *sinful*. Man is now guilty of original sin. We can thus see how once this idea of infinite indebtedness fuses with the bad conscience, the result is gruesome indeed: self-"torment without end" (GM II:22). One now lacerates oneself not only for one's specific misdeeds but also for being a wretched, loathsome creature. One is thus initiated into the ascetic ideal, the main topic of the *Genealogy's* third essay.

What is the psychological resource on which this intensified self-torture draws? Nietzsche explains the phenomenon by citing an excessive "psychical cruelty," a "madness of the will" (GM II:22). How to explain this excess cruelty? Wasn't it all transformed through the process of internalization into non-Christian empirical guilt? The answer I offer is this: the bad conscience did not provide enough of an outlet for *all* of one's cruelty. The more society's strictures and sanctions on freely expressing one's cruelty got increasingly severe, and the more the general populace enjoyed less and less

⁵² Janaway helpfully suggests that moralization involves the idea that it is now considered morally good to feel guilty and torture oneself for one's misdeeds (Janaway 2007, 141–142). There is more at stake here, however.

opportunity to express their cruel instincts freely, for example, by venting them upon neighboring societies (see Ridley 1998, 22), the more there formed a surplus of cruelty in need of discharge. The bad conscience offers a channel for blowing off *some* of the steam but not all of it. Hence, excess internalized cruelty, pressing for release, makes the bad conscience turn *active*. In contrast to the *reactive* self-tortures of bad conscience and guilt over a *specific* misdeed a subject commits, the tendency to think of oneself and perpetually feel oneself to be sinful or corrupt in one's very being is symptomatic of the working of one's *active* instincts. The instincts of cruelty thus gnaw at themselves, not for giving rise to this or that action on this or that occasion (empirical bad conscience) but for existing at all. Christian guilt, the concept of ontological guilt in religious garb, is thus an expression of the *active* bad conscience which is now "spreading within . . . like a polyp" (GM II:21). This is "the *will* of man to find himself guilty and reprehensible to a degree that can never be atoned for"; "this man of the bad conscience has seized upon the presupposition of religion so as to drive his self-torture to its most gruesome pitch of severity and rigor . . . what *bestiality of thought* erupts as soon as he is prevented just a little from being a *beast in deed*!" (GM II:22).⁵³

Reginster (in an unpublished manuscript) raises an important question here: why is this "guilt before God" *guilt* at all? After all, he explains, "The mere possession of animal instincts does not seem to fall within the purview of [the 'sinner's'] free agency, and therefore to be something for which he can hold himself responsible, and therefore blame himself in the manner characteristic of guilt." I believe this puzzle presupposes a conception of guilt that Nietzsche does not share, namely, that "guilt remains *essentially* connected with free will and blameworthiness" (emphasis added). As we saw, bad conscience in its basic

⁵³ For more on the dynamics of cruelty in Christian guilt and how it joins forces with *ressentiment*, see Elgat 2017, chap. 5.

psychological constitution does not involve the idea of free will at all; the thought of free will becomes attached to guilt and is used to justify it only at a much later stage. And though the notion of free will arises roughly at the time of the emergence of Christianity according to Nietzsche (as part of the slave revolt), the transformation of bad conscience to Christian guilt does not presuppose the thought of freedom, either. Self-inflicted cruelty operates here independently of any such presupposition or rational justification; it is, after all, a “*madness of the will*” (GM II:22; emphasis added) that is at work here. That being said, once the idea of free will is used to think about guilt and justify it, the intellectual problem arises for Christian theologians and Christianity-influenced philosophers of how to make sense of this guilt of original sin—the problem that Reginster raises. This is precisely the problem of ontological guilt that Kant, Schelling, and Schopenhauer wrestle with in one way or another, a problem that, as we are now in a position to appreciate, is a complete *pseudo*-problem from Nietzsche’s perspective.

VIII. Concluding Reflections and Criticism

Where does Nietzsche think all of this leaves us? We saw that guilt for Nietzsche is unjustified, for it relies on the fiction of free will and is just a result of the workings of animal psychology. What, then, should the future of guilt be? Nietzsche, it seems, thinks that by knocking down this “ideal,” he is erecting, or at least clearing the ground for the erection of, a new ideal (GM II:24). What ideal, however, does Nietzsche think he is knocking down? Nietzsche’s answer: “*All* ideals hitherto” (GM II:24; emphasis added). Nietzsche writes: “Man has all too long had an ‘evil eye’ for his natural inclinations, so that they have finally become inseparable from his bad conscience.” And he then immediately adds:

An attempt at the reverse would *in itself* be possible—but who is strong enough for it?—that is, to wed the bad conscience to all the *unnatural* inclinations, all those aspirations to the beyond, to that which runs counter to sense, instinct, nature, animal, in short all ideals hitherto, which are one and all hostile to life and ideals that slander the world. . . . Is this even possible today? (GM II:24)

All ideals so far have been all different manifestations of the *ascetic* ideal—an ideal that champions life denial, nature denial, body denial. The task now, Nietzsche suggests without elaborating, is to try to feel guilty precisely for this ideal and its various manifestations: to feel the sting of bad conscience over any thought, judgment, valuation, or action that slanders the body and its instincts.

It should be noted how this practical conclusion differs from the one we saw Nietzsche deduce in his earlier engagement with guilt in HH 39. There Nietzsche claims that the displeasure of guilt “can be given up” once we realize that guilt is unjustified given the absence of free will. Did Nietzsche change his mind? I believe his genealogical analysis itself gave him reason to be skeptical of the possibility of us ever getting rid of the feeling of guilt (as long as we continue to live in society): the assumption of freedom (in the sense of leeway freedom) used to justify guilt (TI, Errors, 7), though false, does not lie at the origin of guilt like a premise to a conclusion—indeed, it does not lie at the origin of guilt at all but was rather adopted at a later stage of historical-psychological development and used to rationalize the assigning of guilt. Consequently, as Rée recognized, guilt would not be avoided once this assumption is removed. This, however, does not change the fact that guilt is unjustified from Nietzsche’s perspective, or it is as justified as the explosion of a steam engine once its safety valves are blocked and no discharge of excess steam is possible.

So should guilt be abandoned or retained? A possible way to combine the two options is the following: Nietzsche’s idea might be not that we should give up the unpleasant feeling of guilt altogether

(what is impossible) but that we should rather discard the moral and religious interpretations of guilt⁵⁴ and utilize the now recognizably *unjustified* feeling by *directing* it at a different object, namely, at the ascetic ideal in all its metastases, for the sake of a this-worldly life affirmation, thus *endowing* it with genuine justification. From this perspective, the problem with guilt is not the negative feeling as such but the values/justifications to which it is harnessed.⁵⁵

Let me now turn to raise a number of concerns about Nietzsche's genealogy of guilt. A central problem for Nietzsche, which I have emphasized, has to do with the workings of the psychological mechanism that is supposed to generate the feeling of guilt subsequent to a misdeed. We saw that Nietzsche thinks that at a relatively early stage, there forms an "intertwinement" of debt and suffering; subjects become conditioned to expect punishment in response to their misdeeds, and while this might plausibly be thought to involve a certain feeling of uneasiness and fear, this account makes no mention of the internalized cruelty that stands at the very basis of guilt for Nietzsche. What, then, is the mechanism that explains how guilt draws on one's internalized cruelty?

Here is a possible answer on Nietzsche's behalf.⁵⁶ While the anticipation of suffering (as punishment) does not in itself amount to guilt, the displeasure it involves triggers—just like any suffering on Nietzsche's view (GM III:15)—a discharge of self-directed *ressentiment*, the instinctive desire to take revenge and inflict suffering upon the perceived cause of one's displeasure, which in this case is identified by the subject as oneself, for, after all, it is recognized that it is because one has done wrong *oneself* that one feels the displeasure of anticipatory punishment. This *ressentiment*, which is in its essence a *reactive* psychological force (GM I 11, II 11), draws

⁵⁴ This, I think, is what he has in mind with the call to restore the "*innocence* [*Unschuld*] of becoming" (TI, Errors, 8).

⁵⁵ It is not clear, however, how this redirection of the bad conscience and its attachment to the ascetic ideal is to be achieved.

⁵⁶ What follows is a summary of the view I articulate in Elgat 2017, chap. 5.

upon active forces such as cruelty—in fact, this self-directed *ressentiment* just *is* cruelty become reactive. As I explained above, active instincts, once their free expression is restrained, are internalized and attached to the mechanics of self-torture; some of them thus lose their active nature and become reactive, and now, in instances of guilt, erupt as a *reaction* to the awareness of one's culpability for a specific misdeed that one has performed.⁵⁷

A related problem arises with respect to Nietzsche's understanding of the nature of the feeling of guilt itself. As we saw, it seems that Nietzsche is of the view that the bad conscience in its very first stages involves the emotions of *sadness* and *fear* (GM II:15). But if sadness and fear constitute the earliest form of the bad conscience, how does it transform into the qualitatively different feeling of guilt? It is possible to construe guilt as a blend of the more basic emotions of fear and sadness (see Prinz 2012, 251). But then what are we to make of Nietzsche's central idea that guilt is the result of the internalization and redirection of *cruelty*? How is cruelty related to fear and sadness? Is Nietzsche saying that guilt is a blend of all these rather different components? Alternatively, does cruelty *supplant* fear and sadness at a more advanced stage? Further compounding the problem is Nietzsche's view that cruelty, as an "instinct of freedom," is an expression of what he calls the will to power (GM II:18). But this seems implausible, for, phenomenologically speaking, guilt is not characterized by the feeling of power, instantiated by the overcoming of a challenge or an adversary.

Indeed, the very foundation of Nietzsche's psychological model seems suspect. Nietzsche assumes that for every person, there is a given quantity of cruelty that, if not discharged outwardly, accumulates inwardly and serves as a reservoir for guilt. This

⁵⁷ One difficulty with this suggestion is that it is not clear whether we do indeed *first* experience a level of discomfort associated with the fearful expectation of punishment and *then* experience guilt (the unleashing of cruelty-turned-*ressentiment* upon ourselves), or whether, upon realizing that we have done something wrong, we immediately start to feel guilt. Indeed, it seems that the latter is more phenomenologically accurate.

picture can be challenged in many ways. One can hold that when cruelty is not discharged, it evaporates or that it *does* get discharged but in micro acts of malice directed at others (which Nietzsche seems to think is true; see HH 50). Alternatively, one can reject the assumption that cruelty can be thought of as constituting a certain fixed quantity (as if it were a gas) or hold that even if a certain quantity of cruelty *is* accumulated and directed inward, it is not guilt that draws on it but other forms of self-torture. And even if we agree that guilt draws on instincts of cruelty that naturally flow outward, guilt would not necessarily require prior inhibition and repression but merely redirection—from being discharged outwardly to inwardly. Possibly, once one learns to think of oneself as another (to see oneself from the outside, so to speak), one can, in guilt, direct one's cruelty at oneself without the prior inhibition and repression Nietzsche describes.

Another deficiency of Nietzsche's genealogy is that it focuses on how the ability to feel guilt was gradually acquired by the species, but it has very little to say about how new members of society come to develop this ability in the process of the development of individual psychology. While it is possible to extract elements from Nietzsche's account that can potentially help us understand child development, his account is phylogenetic in essence and leaves the ontogenetic question almost completely untouched. A reason for this presumably has to do with a Lamarckian strain in Nietzsche's thought: the view that acquired traits are transmitted from one generation to the next. Thus, Nietzsche perhaps was of the view that there is no need to tell an ontogenetic story, since children acquire the capacity for guilt as an heirloom from previous generations; it is because creditors and debtors inflicted pain on each other thousands of years ago that children today possess the relevant mental structure. To the extent that we find this Lamarckism untenable—as we should—we will feel the need to look elsewhere for a satisfactory account. Though Rée, too, focuses his genealogical lens on humanity's prehistory, and though, as we saw, Rée

assumes the possession of some agential “sovereign” capacities which Nietzsche (phylogenetically!) explains, Rée nevertheless emphasizes the manner in which education forms the minds of children and thus helps produce the ability to feel guilt.⁵⁸

Problems arise for Nietzsche’s phylogenetic story as well. As we saw, Nietzsche begins his genealogy of guilt with the creditor-debtor relationship, which is embedded in that of the buyer and seller. There are, however, reasons to doubt the accuracy of this story. Thus, David Graeber criticizes Nietzsche for thinking that buying and selling “precede any other form of human relationships” (Graeber 2011, 76). Graeber claims that here Nietzsche is following in the footsteps of Adam Smith in claiming that economic transactions between individuals come first and that “society comes later” (76)—a “bourgeois” (78) starting point that in Graeber’s view is “insane” (78).⁵⁹

In mitigation of this criticism, it is important to note that Nietzsche does not single out just buying and selling as standing at the basis of human civilization, but he also mentions “barter, trade, and traffic” (GM II:4) as well as “contriving equivalences” and “exchanging” (GM II:8). And indeed, Graeber himself claims that relations of exchange, as well as communal and hierarchical relations, “always exist everywhere” (Graeber 2011, 113) and form the basis for economic relations in a stricter sense (94).⁶⁰ It is thus not clear that Nietzsche and Graeber hold radically divergent views on this score.

⁵⁸ In WS 52, Nietzsche touches upon the way in which education of children provides their conscience with certain specific content—it does not address the development of the ability to experience a guilty conscience itself, however. See also D 34; GS 335.

⁵⁹ Graeber actually holds that Nietzsche knew it was insane but wished to unsettle his bourgeois audience by showing what shocking conclusions follow once you start thinking about human relationships as based fundamentally on relations of trade (Graeber 2011, 78).

⁶⁰ Nietzsche also holds that even in those societies where the preservation of the community was a top priority—an idea reminiscent of the idea of the morality of custom—there was “constant little exercise of consideration, pity, fairness, mildness, reciprocity of assistance” (BGE 201), all elements that fall into what Graeber dubs “communal relations.”

But there is another problem. Graeber also criticizes Nietzsche's view that

those original barbarian law codes that tabulated so much for a ruined eye, so much for a severed finger, were not originally meant to fix rates of monetary compensation for the loss of eyes and fingers, but to establish how much of the debtor's body creditors were allowed to take! Needless to say, [Nietzsche] doesn't provide a scintilla of evidence (none exists). (Graeber 2011, 77)

But Graeber himself gives the example of ancient Egyptian society (Nietzsche gives a different example relating to Egypt in GM II:5), where

when a debtor failed to repay his debt on time, his creditor could take him to court, where the debtor would be required to promise to pay in full by a specific date. As part of this promise—which was under oath—the debtor also pledged to undergo 100 blows and/or repay twice the amount of the original loan if he failed to pay by the date specified. (Graeber 2011, 218, quoting Versteeg 2002)

So Nietzsche is not completely off the mark—there *were* punishments for failure to repay one's debts, punishments directed at the debtor's body. It remains, however, that Nietzsche's gory account is an exaggeration, and to the extent that his genealogy presupposes millennia of corporeal punishment as compensation for failure to repay one's debts, then it stands on shaky grounds.

The mechanism of punishment in Nietzsche's analysis is tethered to actions that, being expressions of the instincts of violence, harm and endanger the community; it is these kinds of actions that one ultimately learns to feel guilty about. Thus, Nietzsche, unlike Rée, cannot explain guilt for not showing compassion or sympathy, for failing to love one's neighbor (where such failures are not motivated

by the “instinct of freedom”).⁶¹ Nietzsche can fix this issue if he gives up on the notion that bad conscience involves cruelty targeting *itself* and if he extends the practice of punishment so that it also covers failures of beneficence of various kinds; the internalized instincts of cruelty, when their host fails to act compassionately toward others (as the community might expect in certain circumstances), would then lacerate him or her precisely for this failure. Nietzsche would still, however, be unable to account for the mixed phenomenon Rée mentions where an agent feels both guilt and the pain of frustrated compassion, given that the latter, it seems, is missing from his inventory of animal instincts (at least, in the *Genealogy* and his later period of thought). I will raise one final criticism of Nietzsche’s view in the book’s conclusion. First, though, I turn to examine Heidegger’s thoughts on guilt in his *Being and Time*.

⁶¹ One can be in certain situations blind to the needs of others without this blindness being motivated by cruelty and so on. A strong indication of Nietzsche’s limitations here is given in his claim about “the morbid softening and moralization through which the animal ‘man’ finally learns to be ashamed of *all* his instincts” (GM II:7; emphasis added). But what about instinctive beneficence, instinctive compassion?

6

Heidegger

Being-Guilty as a Condition of Possibility of Guilt

Introduction

As we saw, on the one hand, for the metaphysical tradition, guilt is justified because the guilty agent is a *causa sui*. On the other hand, for the naturalistic tradition, because there is no freedom in the world of appearances and the notion of *causa sui* is philosophically repugnant, guilt is never justified, though it is explainable. We seem to have before us a contradiction of thesis and antithesis. Is there a way to move beyond it?¹

In this chapter, I argue that Martin Heidegger can be read as providing us with a synthesis of sorts of the two traditions. Specifically, I focus on Heidegger's analysis of Being-guilty (*Schuldigsein*) in his 1927 *Being and Time* (Division II, Chapter II)² and argue that Heidegger's approach holds the promise of showing how guilt is possible without becoming trapped in the apparent dead end of the metaphysical tradition on the one hand or the psychological reductionism of the naturalistic view on the other. Thus, the argument will be developed that in Heidegger's view, while we are indeed not *causa sui*, as the naturalists hold, we are nevertheless guilty as such or are characterized by ontological guilt, as the metaphysicians

¹ Putting aside, of course, compatibilism about free will, which all of the philosophers discussed in this book reject or do not consider as possibly justifying guilt.

² Henceforth BT. All references, unless otherwise noted, are to the English version's pagination of Macquarrie and Robinson's English translation of Heidegger's *Being and Time* (Heidegger 1962).

hold, and this is precisely because for Heidegger, *not* being *causa sui* is a condition of our ontological guilt, our Being-guilty. Moreover—and here again, Heidegger can be seen to resemble the metaphysical tradition—it is our Being-guilty that makes our factual or empirical guilt possible. It is thus on the basis of his novel philosophical approach—his existential phenomenology of the human being or *Dasein*—which eschews both the conceptual framework of the metaphysical tradition as well as the naturalistic approach to man, that Heidegger can be seen to provide us with a conception of guilt that avoids the difficulties that beset both traditions and yet retains some of their merits and insights.

While I cannot hope here to introduce the reader to *Being and Time* and its challenging conceptual apparatus,³ let me nevertheless provide the following ultra-condensed presentation of some of his central ideas before I move on to focus on his analysis of guilt, where I will fill out some of the details in Heidegger's picture as needed.

Heidegger's fundamental aim in *Being and Time* is to reopen for philosophical investigation the "question of Being" (e.g., BT 21), which is, roughly, the transcendental question about what makes possible the different ways of understanding the different kinds of Being or ways-to-be of entities. To make headway into this question, Heidegger chooses to focus on a specific kind of entity, namely, the human being or *Dasein*, for this entity, according to him, possesses an implicit understanding of Being, that is, of the different ways of Being that different entities can have, including its own Being. To interrogate this entity, *Dasein*, Heidegger employs in the first division of *Being and Time* a phenomenological approach that attempts to describe and lay out the kind of Being that *Dasein* has. Heidegger calls this kind of Being "existence" (e.g., BT 67). *Dasein*'s Being, in contradistinction to the kind of Being of mere objects or tools, is interpreted by Heidegger as "Being-in-the-world" (78ff.).

³ For a couple of valuable introductions, see Dreyfus 1993; Polt 1997.

Dasein's Being is such that it always already finds itself in what Heidegger calls its "there" (172ff.). It finds itself immersed in a world of meaning and significance which is not of its own making, a world to which its relation is first and foremost practical, for Dasein is typically engaged in some practice or other that makes sense to it in terms of its own practical self-understanding. This practical know-how is referred to by Heidegger with the term "understanding" (182ff.). In addition, the "there" involves being in a "state-of-mind" (172ff.), a form of attunement in terms of which things show up for Dasein as mattering in some way or other. Finally, the "there" is articulated in terms of what Heidegger calls "talk" or "discourse" (203). Furthermore, Dasein's Being-in-the-world is characterized by what Heidegger calls "Being-with" (e.g., 153ff.): Dasein is the kind of entity that essentially finds itself in the world alongside others (other Daseins) and understands itself in terms of its relations to those others. At the heart of Dasein's Being is the fact that its own Being is an issue for it (e.g., 236); Dasein is the kind of Being that is characterized by the peculiarity that the question of its own Being as Dasein in general and as a particular Dasein is something that Dasein attempts to settle, and it is in terms of the world in which it finds itself that it thus attempts to make sense of itself. At a more fundamental level, Heidegger explains, Dasein's "Being-in-the-world is essentially care" (237), where "care" encompasses all the various structures of Dasein's constitution. Dasein *cares* about its Being, and this care expresses itself, we can say, in the various ways in which it is in the world. It is in terms of these features that Dasein is also said to *disclose* Being (e.g., 408), that is, again, roughly, to have an understanding access to it on the basis of which it can comport itself toward specific entities (tools with the kind of Being of the "ready-to-hand," or mere objects, with the kind of Being of the "present-at-hand"). All of this is presented (in much greater detail, of course) in the first division of the text. In the second division, the attempt is made to grasp Dasein more "primordially" (e.g., 274) and "as a whole" (e.g., 276): while in the

first division, Dasein is mostly interpreted in its *inauthentic* or *un-owned* (*uneigentlich*) form, that is, as lost in the anonymous and public way of interpreting the world—which Heidegger calls the “they”—the task now is to reveal its potentiality for Being *authentic* or *owned* (*eigentlich*). It is in this context that Heidegger goes on to discuss Dasein’s Being-guilty. What *is* this Being-guilty, however? How is it conditioned by Dasein’s *not* being *causa sui*? Indeed, in what sense *isn’t* Dasein *causa sui*? And how does Being-guilty make empirical guilt possible?

After introducing some of the main concepts and themes of Heidegger’s discussion of guilt, I turn to reconstruct Heidegger’s transcendental argument to the effect that our Being-guilty is a necessary condition of the possibility of factual guilt, that is, guilt for our specific misdeeds. I then turn to discuss Heidegger’s concept of the call of conscience and then, in the final section, examine his concept of wanting-to-have-a-conscience.

I. Being-Guilty: A Summary of Heidegger’s Central Ideas

That Heidegger was familiar with Kant’s, Schopenhauer’s, and Nietzsche’s conceptions of conscience and guilt is clear (BT 495, footnote vi). It is thus safe to assume that he was also familiar with Nietzsche’s criticism of these earlier conceptions of guilt—specifically with Nietzsche’s criticism of the problematic notion of *causa sui*. It is therefore legitimate to try to read Heidegger’s analysis of guilt in *Being and Time* as involving (among other things) an attempt to answer Nietzsche’s criticism.

In the second chapter of the second division, “Dasein’s Attestation of an Authentic Potentiality-for-Being, and Resoluteness,” Heidegger provides an “ontologically adequate Interpretation of the conscience” (319) with the aim of establishing whether, in the call of conscience, “an authentic potentiality-for-Being of Dasein”

is “attested” (312). This is important, for it constitutes part of Heidegger’s attempt to bring Dasein’s Being-a-whole into grasp so as to attain a more ontologically secure ground on which Heidegger can further investigate the temporal character of care and, ultimately, the meaning of Being (I am here, of course, condensing several difficult lines of thought).

Let me summarize the main elements of Heidegger’s discussion in this second chapter. Heidegger first provides an existential-ontological analysis of conscience that relates the conscience to the “most universal structures of state-of-mind, understanding, discourse and falling” that make up the “*Being of the ‘there’* as disclosedness” (315). Accordingly, Heidegger’s phenomenological findings are that in conscience, Dasein calls its *own Self* (317) in its lostness in the “They” into which it has fallen (322) and that this calling—though it strictly speaking says nothing (318)—is a “mode of discourse” (316). Moreover, in the call of conscience, Dasein, as caller, “*finds itself in the very depth of its uncanniness*” (321) and “has been individualized down to itself in its uncanniness” (322), where “[u]ncanniness reveals itself authentically in the basic state-of-mind of anxiety” (321). Finally, the call, though it “does not report events” and “calls without uttering anything” (322), gives Dasein something to understand, namely, that Dasein is “Guilty” (e.g., 325)—it calls Dasein *to* its ownmost Being-guilty which “remains closed off from the they-self” (334). This shows, Heidegger claims, that “The call of conscience . . . has its ontological possibility in the fact that Dasein, in the very basis of its Being, is care” (323).

But how should we understand this primordial Being-guilty? How can we distinguish it from the ordinary, everyday understanding of guilt? Heidegger analyzes the ordinary significations of Being-guilty (i.e., the ordinary significations of what it is to be guilty) and claims they involve two basic components that come together in the following formal conception of the everyday conception of guilt: “Being-the-basis for a lack of something in the Dasein of an Other, and in such a manner that this very Being-the-basis

determines itself as “lacking in some way” in terms of that for which it is the basis” (328). In other words, as a result of being the basis of a lack in another, one incurs a lack in oneself, the nature of which refers to the lack in the other. And Heidegger immediately clarifies: “This kind of lacking is a failure to satisfy some requirement which applies to one’s existent Being with Others,” one example of which involves “the breach of a ‘moral requirement’ [*sittlichen Forderung*]” (328) which can lead to the “Other’s becoming endangered in his existence, led astray, or even ruined” (327). For example, when one steals from another (and thus fails to abide by a central moral principle), thus causing a lack in the other (the other will now miss the thing stolen or will suffer the consequences of the theft), one is characterized by a lack, a moral failure, the content of which is determined by that which one has done, a specific act of theft.

Heidegger, however, explains that this definition cannot be *ontologically* satisfactory, that is, it cannot grasp the fundamental nature of the phenomenon of guilt in relation to Dasein, for at least a couple of reasons. First, it makes reference to the concrete debts to others that one might incur. But this element, which has to do with our “concernful Being with Others” (328), characterizes the “ordinary phenomenon of ‘guilt’” where Dasein is concerned with “reckoning up claims and balancing them off” (328 and therefore cannot guide us toward an understanding of the *primordial* sense of Being-guilty that belongs to Dasein’s very being. Consequently, the reference to these ordinary, everyday instances of guilt must “drop out” (328).⁴ Second, the definition includes elements that cannot properly apply to Dasein; specifically, “here too ‘guilt’ is still necessarily defined as a *lack*—when something which ought to be and

⁴ I have thus revised—thanks to comments by an anonymous reader—my prior understanding (Elgat 2020) of this passage (BT 328): in contrast to my previous reading, Heidegger is *not* saying here that the primordial concept of Being-guilty should exclude *all* reference to others but only that it must exclude reference to the everyday or ordinary Being with others which involves the reckoning up of specific claims.

which *can* be is missing. To be missing, however, means not-Being-present-at-hand" (328). But since Dasein's Being is not that of the present-at-hand, the formal everyday definition of Being-guilty cannot possibly apply to it in its most primordial Being.⁵

Nevertheless, Heidegger claims that Being-guilty in the primordial, ontological sense *does* retain something of the character of the "not" (329) that is involved in the formal definition of the everyday conception of guilt and adds that, like the latter, the primordial sense also includes the idea of "Being-the-basis for" (329). He therefore reaches the following formal existential idea of the "Guilty!": "Being-the-basis for a Being which has been defined by a 'not'—that is to say . . . *Being-the-basis of a nullity*" (329). After some clarification, Heidegger reaches a fuller definition of Dasein's Being-guilty: "Being the basis of a nullity (and this Being-the-basis is itself null)" (331), or, more compactly, "the null *Being-the-basis of a nullity*" (353). How to understand these two "nullities"? I will introduce them now very briefly and will return to them in more detail.

Let's start with the nullity of which Dasein is the basis. Heidegger explains this by saying:

what we have here is rather something existentially constitutive for the structure of the Being of projection. The nullity we have in mind belongs to Dasein's Being-free for its existentiell possibilities. Freedom, however, *is* only in the choice of *one* possibility—that is, in tolerating one's not having chosen the others and one's not being able to choose them. (331)

⁵ A concern arises here: it is not clear why the lack which is involved in the formal concept of the everyday notion of Being-guilty should be understood in terms of the Being of the present-at-hand. Specifically, when one owes something to another and is thus responsible for a "lack" in that other person, this lack cannot be understood as substance in the Aristotelian or Cartesian sense, it cannot be understood merely in terms of its physical properties, and it cannot be understood in a manner that is absolutely extricated from the world and its normativity-laden involvements. Yet these are three central ways of understanding the category of the present-at-hand (cf. Golob 2014, 16–17).

Turning now to the other nullity, how to understand the claim that Dasein is a null basis? Heidegger claims that “being a basis . . . [is] existing as thrown” (330).⁶ This means, first, that we should understand being a basis in relation to the concept of thrownness. Heidegger glosses Dasein’s thrownness as “it [having] been brought into its ‘there,’ but *not* of its own accord” (329). The “there,” in turn, is constituted by a state of mind, understanding, and discourse (171–172). Differently put, to be thrown, to be a basis, means precisely *not* to be in control over the various factors that make up the being of the “there” but rather, as he puts it, to be in such a way that “Dasein constantly lags behind its possibilities” (330). Heidegger claims earlier that “‘thrownness’ is meant to suggest the *facticity* of its being delivered over [Überantwortung]” (174). Dasein is thus not responsible—*verantwortlich*—for its Being but is rather, passively, delivered over to itself, and thus lacks ultimate responsibility for being at all and for the kind of Being that one is (Dasein, rather than, say, a stone).⁷ We thus arrive at the idea that “‘Being-a-basis’ means *never* to have power over one’s ownmost Being from the ground up. This ‘*not*’ belongs to the existential meaning of ‘thrownness’”—“Dasein is not itself the basis of its Being” (330).

What all of this means, I want to suggest, is precisely that Dasein is *not causa sui*; it does not create itself from the ground up but is rather always already thrown into a “there” that is not of its own

⁶ Steven Crowell mentions several ways that have been suggested by interpreters for how to understand this “basis” or “ground” (Crowell 2007, 56), but given that Heidegger clearly glosses “being a basis” as ‘existing as thrown’ (BT 330), I cannot share the interpreters’ uncertainty. Why, then, use the language of “ground” or “basis” instead of talking directly about “thrownness”? I believe the answer is twofold. First, Heidegger, by using the concept of “basis,” preserves continuity with his definition of the everyday conception of guilt which also employs the notion of basis. Second, the notion is useful to underscore precisely how Dasein is *not* its own doing, that is, can never *ground* itself, or, on my reading, cannot be *causa sui*. I will return to this.

⁷ Mark Basil Tanzer holds that our ontological guilt is a “lack or deficiency . . . for which Dasein is somehow responsible” (Tanzer 2002, 72). If this is meant to suggest that Dasein is somehow the origin of this lack, then this stands in contrast with the idea that we are *given over to our Being* and are not responsible for it and the “not” that characterizes it as such.

making but that, nevertheless, “as existing, it must take over” (330). This rules out the metaphysical notion that we are factually guilty because we are responsible and guilty for ontologically constituting our intelligible character, that is, for our ontological guilt. To be responsible in this latter sense necessarily presupposes the rejection of our thrownness, for it means that we somehow prepare ourselves behind the scenes, so to speak, *before* we step on the stage; it thus necessarily presupposes a thoroughly flawed understanding of the Being that we are, according to Heidegger.⁸

Does this then mean that for Heidegger, we cannot be factually guilty, that is, guilty for our specific misdeeds? Not at all. While Heidegger rejects the idea that we are *causa sui*, he nevertheless maintains that we are ontologically, essentially Being-guilty and that this is precisely *a necessary condition of the possibility of factual guilt*. He thus shares with the metaphysicians the idea that we are guilty as such, and moreover, he shares with them the idea that it is this guilt that makes our factual guilt possible. In other words, he agrees with them that our essential guilt is a transcendental condition for our factual guilt in its various forms, moral and otherwise. Thus, Heidegger claims in transcendental fashion that “*Being-guilty does not first result from an indebtedness [Verschuldung], but that, on the contrary, indebtedness becomes possible only ‘on the basis’ of a primordial Being-guilty.*” (329; emphasis on “only” added). And again:

Not only can entities whose Being is care load themselves with factual guilt, but they *are* guilty in the very basis of their Being;

⁸ A second, related problem with the metaphysical tradition from Heidegger’s perspective is this. The Kantian or Schopenhauerian subjects, in constituting or creating themselves, to be guilty of this ontological guilt must be seen as understandingly appealing to certain norms or values on the basis of which they create themselves as good or evil. But this assumes that the self, in abstraction from the world, that is, the noumenal self, has access to such norms and values. But this goes against Heidegger’s view that all normativity is to be found in the world into which Dasein is thrown. See Golob 2014, e.g., 230–231.

and this Being-guilty is what provides, above all, the ontological condition for Dasein's ability to come to *owe* anything in factually existing . . . [and for] the possibility of the "morally" good and for that of the "morally" evil—that is, for morality in general and for the possible forms that this may take factually. The primordial "Being-guilty" cannot be defined by morality, since morality already presupposes it for itself. (332)

We can thus see that for Heidegger, Being-guilty is a condition of possibility for factual guilt. Moreover, Being-guilty is a *necessary* condition for factual guilt: the latter becomes possible *only* on the basis of Being-guilty. And since we are not *causa sui*, for Heidegger, unlike Kant, Schelling, and Schopenhauer, we are not guilty for this ontological guilt—we are *delivered over* to it.

The last sentence in the quotation just given implies another important point, namely, that the metaphysical idea that subjects can become factually guilty by loading themselves with guilt on the ontological level is mistaken not only in its presupposition of the subject as *causa sui* but also in that it presupposes a conception of primordial guilt that is defined by morality. But this would be to think of Dasein as possessing access to, and as defined by, certain moral norms even in abstraction from the world into which Dasein is thrown—which Heidegger rejects. It would, moreover, privilege morality as the exclusive domain of ontological guilt, while Heidegger is of the view that Being-guilty as such has nothing to do specifically with morality. In other words—and this is an important point—Being-guilty, while referring to ontological guilt, a guilt that belongs to the human being as such, is utterly amoral: it does not refer to any kind of moral deficiency, sin, or depravity of the human being.

Dasein, then, is both Being-guilty and not *causa sui*. As I will argue in detail, for Heidegger, it is not that we are essentially guilty and in addition just happen to not be *causa sui*. The claim is stronger: we are guilty in our Being precisely *because* we are not

causa sui. Differently put, our ontological guilt consists precisely in that we never have power over ourselves “from the ground up” and always “lag behind” our possibilities.

But how are we to make sense of these transcendental claims? In what precise way does Dasein’s ontological Being-guilty—Dasein’s being a null-basis of a nullity—make its factual guilt possible? What is Heidegger’s argument? Moreover, how precisely is it that Dasein’s *not* being *causa sui* constitutes our Being-guilty? And how are we to make sense of Dasein’s “nullities”—how are they related to Dasein’s ownmost Being-guilty? While the last several years have witnessed growing scholarly attention to the central themes and concepts of the second division of *Being and Time*,⁹ no detailed attempt to reconstruct the argument has been made in anglophone Heidegger scholarship. I will now turn to address these questions.

II. Being-Guilty as Being Responsible

Before I proceed to present my own interpretation, I will address a possible way of thinking about Heidegger’s concept of guilt and his transcendental argument. Specifically, it seems that one can take the following (very fast) transcendental line. It can be held—in line with an interpretation of the term suggested by John Haugeland (2013, e.g., 153, 207)—that ontological guilt *is* responsibility: Being-guilty is Being-responsible,¹⁰ and the human being as such is a *locus of responsibility* or *accountability*.¹¹ But—the argument can be developed—if we are responsible ontologically, then we are also factually responsible for specific deeds that we perform

⁹ See McManus 2015 for a collection of articles focusing on, among other topics, death, guilt, and authenticity.

¹⁰ See also Pattison (2015, 59) for an interpretation that strongly ties Being-guilty to responsibility.

¹¹ This responsibility is to be distinguished from the responsibility that involves *responsiveness* to certain norms which Haugeland calls “routine responsibility” (Haugeland 2013, 238–239, also 204–205).

or fail to perform (other things being equal). But being factually responsible for a misdeed is to meet a necessary condition for factual guilt given that we cannot justifiably be held guilty or experience guilt for a misdeed where responsibility for the misdeed is lacking. Hence, Being-guilty/responsibility makes factual guilt possible—it is a necessary condition of it.

It is not my intention to offer here a detailed reconstruction and examination of Haugeland's interpretation of the notion of responsibility in the context of his work on Heidegger, so I will limit myself to the following critical observations.¹² First, and quite simply, while without a doubt ingenious, Haugeland's suggestion that we read Heidegger's "Being-guilty" as "responsible" is not supported by direct or explicit textual evidence;¹³ Heidegger never, to my knowledge, explains Being-guilty in *Being and Time* as responsibility or accountability.¹⁴ Second, it is important to remember that guilt and responsibility, though admittedly very closely related, are *different* phenomena, and Haugeland's interpretation assimilates the two. In contrast, the interpretation offered here takes Heidegger's talk about *guilt*—and *not* about responsibility—seriously and strives to understand in what sense our ontological guilt, though different from our factual guilt, is nevertheless a kind of *guilt*.

Third, we should ask, what notion of responsibility is invoked here in Haugeland's notion of Being-responsible? How to understand the agent's ontological responsibility?¹⁵ Haugeland claims

¹² It is worth remarking that I have revised my views on Haugeland's reading since my 2020 paper on Heidegger and guilt (Elgat 2020).

¹³ It should be kept in mind, however, that already in an early essay, Haugeland claims that his reading of *Being and Time* is "non-standard" and "free-wheeling" (Haugeland 2013, 3).

¹⁴ As we shall see below, to become responsible (*verantwortlich*) in terms of a *specific* potentiality-for-Being that one has chosen, that is, as a teacher or a nurse (i.e., existentially, rather than existentially), *presupposes* for Heidegger what he calls "wanting to have a conscience" or projecting oneself upon one's Being-guilty (BT 334).

¹⁵ I write the "agent" rather than "Dasein" because in Haugeland's rather idiosyncratic interpretation of Heidegger, the term "Dasein" refers not to a specific person but to something like a "way of life" (e.g., Haugeland 2013, 31). This makes his interpretation of Being-guilty all the more problematic given that for Heidegger, it is Dasein that is

that the responsibility at issue is responsibility for one's "whole life," which the authentic, resolute agent must take over. It's a responsibility for oneself "as a whole" and for "who" one is (Haugeland 2013, 208).¹⁶ Now, the history of philosophy provides us with at least two ways of thinking about responsibility in a way that can ground responsibility or accountability for a specific action and consequently can ground factual guilt (in case the action performed was wrong in some sense), a responsibility that presupposes the ability to act otherwise (dependent on the notion of libertarian, leeway freedom) and a compatibilistic notion of responsibility which can be cashed out in various ways (e.g., Wallace 1994; Fischer and Ravizza 1998). But how can the responsibility that Haugeland identifies at the root of the human agent—responsibility for *who* one is as a whole—similarly ground factual guilt? Illuminating and suggestive as Haugeland's interpretation might be, this is not immediately clear.¹⁷

Fourth, and relatedly—and this is especially important for the interpretation I am developing in the book as a whole—a central point of my analysis in the preceding chapters was that responsibility in the substantial sense, a central condition of guilt according to the figures examined, is in the last analysis *unjustified* (reliant as it is on the notion of *causa sui*) and that *therefore* guilt is unjustified as well. Consequently, simply *positing* our ontological responsibility in this way, that is, by reading Being-guilty as responsibility in a sense that can potentially ground factual guilt, runs the risk

Being-guilty, but to ascribe guilt or even responsibility to a way of life is rather puzzling, to say the least. I will put this difficulty aside.

¹⁶ In another instance, while providing commentary on Heidegger's "*zu sein*," Haugeland claims that Dasein has its being as "its *burden* or *responsibility*" (Haugeland 2013, 85). Responsibility as a burden or, we can say, a *task* comes close to the idea of Being-guilty which I will develop.

¹⁷ I am claiming here not that the explanation required cannot in principle be provided but that Haugeland himself does not take up this specific question. In general, his interpretation is mostly focused on the agent's responsibility for *specific* disclosures of Being, such as those in the various sciences, rather than with the agent's responsibility for being itself as a whole (Haugeland 2013, e.g., 212–220).

of begging the question against an entire chapter in the history of philosophy. On the other hand, to read Heidegger as saying that we are fundamentally ontologically *guilty* (as I will do) is *not* similarly question begging because the conception of ontological guilt I will developed does *not* presuppose any notion of responsibility (or freedom) challenged by the tradition examined in the preceding chapters.¹⁸

Another attempt to think of Being-guilty as tightly connected to the concept of responsibility is that of François Raffoul. Raffoul's approach, in general, suffers from some of the problems just raised, as well as from other issues,¹⁹ but he might claim that the concept of responsibility at stake in *his* analysis of Being-guilty is *not* that of accountability at all: "There is a sense of responsibility which does *not* amount to accountability . . . but will signify instead the very movement of a radically finite existence having to come to itself, and to itself as other, from an inappropriable (and thus always 'other') ground" (Raffoul 2010, 268–269). As we shall see, I am sympathetic to the idea of Dasein coming to itself, but why call this

¹⁸ It might be replied that the notion of ontological responsibility Haugeland offers is also unlike any of the notions criticized or analyzed in the traditions examined, for it is not a responsibility based on leeway freedom or on the possession of certain agential capacities such as self-control. But this raises the question of what, then, constitutes this *ur-responsibility*. Into what elements could it be analyzed?

¹⁹ Raffoul ties the concept of responsibility very closely to the concept of Dasein as such, first by claiming that "one could go as far as to say that the very concept of Dasein *means* to be a responsibility" (Raffoul 2010, 244) and then by claiming that "being-guilty [which Dasein constitutively is] . . . should be understood in the sense of an *archi-ethical* [i.e., ontological] responsibility" (261). However, it is not clear why one should talk about responsibility here at all. Heidegger himself is careful not to use this term and, as we shall see, holds that Being-guilty is rather a *condition* for responsibility (*Verantwortlichkeit*), rather than *being* responsible. It, of course, does not help that Raffoul insists, to support his case, on translating Heidegger's *überantwortet*—delivered over—as "responsible" (*verantwortlich*). Not only is this a problematic translation, but it works to obscure Heidegger's idea, namely, that we are passive with respect to our thrownness. Second, the idea that Being-guilty is responsibility clashes with Raffoul's own claims when he says that "The call of conscience thus calls me back from the disburdened (de-responsibilized) existence in the everyday, back to my own Being-guilty" (265), for this seems to entail that it is possible for Dasein to be *not* responsible (or be less responsible?) in some sense, but since Being-guilty belongs to Dasein as such, this would imply that Dasein can stop being Dasein (or be slightly less Dasein?).

“movement” responsibility? What does this concept share with the notion of responsibility so that it merits the same appellation—“responsibility”? Second, if responsibility in Raffoul’s sense is distinct from accountability, how precisely can it serve to ground guilt in the factual sense, a sense that *is* grounded in the notion of responsibility as accountability? Raffoul, in my view, does not answer these questions.

III. Being-Guilty as a Condition of Possibility of Factual Guilt

a. Ontic Guilt versus Factual Guilt

I now turn to present a different interpretation of Being-guilty,²⁰ in order to reconstruct Heidegger’s transcendental argument and explain how Being-guilty can make factual guilt possible. To answer this question, to bridge the concept of ontological guilt (Dasein’s Being-guilty, its null Being-the-basis of a nullity) and factual guilt (guilt for specific misdeeds), I supply an analysis of an idea that is not explicitly present in *Being and Time* and yet is one that I think is necessary in order to make plausible the transcendental argument. The idea I will introduce and discuss, specifically, is the idea of *ontic guilt*. I therefore rely in the following on a threefold distinction of the ontological, the ontic, and the factual.²¹

²⁰ I was here inspired by Tanzer’s reading, though I nevertheless remain critical of some of his specific formulations. Tanzer claims that Dasein “is guilty insofar as it is required to achieve itself (since it is its own end), but due to the indeterminate structure of the end in itself, it is ineluctably deficient; and so it is called upon to complete an unachievable task—the task of self-achievement, of taking over its death” (Tanzer 2002, 73). While I do agree that Dasein can never attain self-coincidence and that this is part of Dasein’s ontological guilt, I don’t think Dasein is “required” or “called” to achieve itself in this way in some fashion that will inevitably result in failure. Put differently, I don’t think Dasein is a “useless passion,” in Sartre’s phrase.

²¹ My usage of the term “ontic” does not map exactly onto Heidegger’s, but this should not be an obstacle, since I will explain the way I use the term.

The idea of *ontological* or existential guilt is one we are already roughly familiar with: Dasein is guilty in its very Being. I will return to this idea in detail. The concept of *factual* guilt is also one we are acquainted with, but I will now formulate it in terms more amenable to Heidegger's discussion: factual guilt is concerned with not abiding by a specific normative demand that applies to a specific Dasein (and is understood to so apply) in a specific situation. For example, a student might be and feel guilty for not doing her homework on a certain day. In its perhaps most salient manifestation, factual guilt is specifically *moral* guilt—the failure to abide by a specific moral demand that applies to oneself (and is understood to so apply) on a particular occasion.

What, however, do I mean by *ontic* guilt (or *existentiell* guilt)? The concept of ontic guilt concerns the guilt that characterizes Dasein as a being that projects itself onto one or another *possibility of Being*. While ontological guilt abstracts from any content, and factual guilt concerns a particular misdeed, ontic guilt is more determinate than ontological guilt but less specific than factual guilt; by means of this concept, the gap between the idea of Being-guilty and factual guilt will be bridged. Let me explain this concept in more detail.

The basic idea of ontic guilt is that Dasein, as existing, must project itself upon a possibility or possibilities of Being of some kind or another. This projection belongs to what it is to exist and is thus constitutive of Dasein's very Being (BT 185).²² Thus, for example, Dasein can project itself upon the possibility of Being a student or a policeman. In addition, Dasein can, and typically does, "simultaneously" project itself into several possibilities of Being. Thus, one can

²² There is a distinction in Heidegger, which in my view Heidegger himself does not always observe, between potentiality-for-Being and possibility of Being. Roughly, the former indicates the general ability of Dasein to be this or that and characterizes Dasein at the most basic ontological level, while the latter refers to this ability as it is concretized in the specific practical identities that Dasein takes up in its world and characterizes Dasein at what I call the ontic level. A specific action Dasein performs belongs to the factual level.

be a teacher and a parent and a Protestant Christian. But to project oneself in this way, Heidegger claims, is for Dasein to project itself upon a “for-the-sake-of-which” (185), that is, upon a practical identity (Korsgaard 1996) that organizes one’s existence as this or that and that one *cares* about, projection being an aspect of Heidegger’s concept of care. In other words, to project oneself onto a possibility of Being is (partly) to interpret oneself and so conduct oneself *as* this or that, for example, *as* a student or policeman. But, I wish to suggest, to interpret oneself in this way is to consider oneself as *governed by norms*.²³ This means that to *be* a certain possibility of Being necessarily involves seeing oneself as governed by the norms that constitute it—norms that involve duties, commitments, expectations, and so on, of various kinds—norms, moreover, that typically concern one’s relations to certain others within the framework of what Heidegger calls “solicitude” (*Fürsorge*; e.g., BT 157) and—importantly—that one also *cares* about (insofar as one cares about one’s possibility of Being). In addition, since understanding is crucially practical for Heidegger, to understand oneself as an X (father, say) is to possess the ability to *act* on these norms. We might call these constitutive norms *ontic norms* or *ontic demands*.

Now, if one does not possess some understanding of the ontic norms that govern oneself qua X, then one cannot really comport oneself in light of those norms, and if one does not care about them, then one *will* not tend to comport oneself in this way, and, finally, if one cannot act in light of them, then one will not *succeed* in being X. Under such conditions, one cannot be said genuinely to *be* an X. For example, to be a parent (in the nonbiological sense) is to have some sort of understanding of what it means to be a parent (or a good parent), that is, of what one should do and avoid doing as a

²³ I admit I am not fully happy with the concept of norms, for it makes it sound as if what I have in mind is something social and publicly enforced in one way or another, which would make it too tied up with the concept of the “they” (e.g., BT 334), and yet it is possible according to Heidegger to recognize one’s guilt in a manner that breaks with the “they” and its understanding of how one should be.

parent, and to try to act in light of these norms insofar as one cares about one's Being-a-parent. A person who has no understanding of what it means to be a parent, or does not care and thus does not act accordingly, or cares but consistently fails to meet the norms that govern parenthood, cannot really be said to *be* a parent (perhaps, in the latter case, merely an *aspiring* or a *struggling* parent).

Now, what is important about ontic norms is that they apply to Dasein qua X (student, parent, etc.) without necessarily making upon Dasein at any given time or place any specific demand. Thus, one can be a parent without constantly occupying oneself with the discharge of the various obligations or duties that are involved in this possibility of being. While one is at work, for example, one is typically concerned with other tasks than those that make up Being-a-parent. Since for the parent at work, there is presumably no *specific* parent duty or demand that applies to the parent at that specific time, there is no possibility for the parent to acquire and experience *factual* guilt at that specific time with respect to his Being-a-parent. On the other hand, if the parent, say, is supposed to leave work early on a certain day in order to attend a parent-teacher conference, then we could say that there *is* some *factual* demand that applies to the parent at that time, one which, were the parent not to fulfill it, could burden the parent with factual guilt which he could then go on to experience once he becomes aware of his failure.

But, while there might be for the parent no specific parent-related factual demand that applies to him at a specific time, insofar as this person *is* a parent, there apply to him demands of a different kind—demands that constitute his being a parent. This is because to be an X—whatever X is—is to exist under demands (or norms) that do not constantly generate for Dasein specific demands but rather apply to Dasein's being an X *as long as* Dasein *is* an X. Put differently, to exist as an X is to comport oneself in a normative space that is constitutive of this possibility of Being and yields specific duties or demands that apply to oneself on specific occasions (to attend the parent-teacher conference this evening, say).

But not only do these general ontic demands apply throughout as long as one is an X; the *content* of the demands is indefinite. As Heidegger claims, an “indefiniteness [*Unbestimmtheit*] ... [is] ... characteristic of every potentiality-for-Being into which Dasein has been factually thrown” (BT 345). The indefiniteness of the norms that govern a possibility of Being has to do mainly with the following four features which I will touch upon briefly. First, one cannot know in advance when and where an ontic norm will come into effect and yield a specific requirement. Second, it is impossible to formulate the ontic norms explicitly, fully, and precisely. For example, it is impossible, despite the tendency of the “they” to turn these into “manipulable rules and public norms” (334), to spell out with precision and completeness the norms that constitute being a friend.²⁴ Third, the norms are open-ended and subject to change; they are in principle mutable and can either shrink or expand in their field of governance. Fourth, whether and how a norm applies to a specific situation is not always absolutely clear and requires judgment. Nevertheless, importantly, the ontic demands, while not fully definite, are not utterly indeterminate but consist of certain broadly defined and understood norms that govern the possibility of being X, norms that, as already mentioned, continuously apply to Dasein as long as Dasein interprets itself accordingly. This indefiniteness, however, becomes more fully determined, that is, acquires a more definite content in specific circumstances and generates a factual demand.

For example, doing one's homework on a Wednesday evening could meet a concrete factual demand that a student might face at that specific time. But to *be* a student is to be under broader, *ontic demands*. Thus, even if one does one's homework *this evening* and thus discharges a concrete obligation, one still ought to do so on multiple, yet-to-be-determined future occasions. In addition, one

²⁴ See McDowell (1998), where the idea of the uncodifiability of virtue is elaborated. I do not, however, share all the details of McDowell's view.

ought to study for all future tests, not miss school too many times, be respectful and kind toward one's peers and teachers, and so on. In addition, being a student might involve certain demands that one cannot exactly foresee or fully articulate to oneself in advance. To *be* a student is to *be governed* in one's existence by all these various ontic norms. It is to live one's life in a certain manner which is shaped by the self-understanding one takes upon oneself as this or that. Hubert Dreyfus makes a similar point when he claims that "man's ultimate concern is not just to achieve some goal which is the end of a series [e.g., discharge a specific factual duty]; rather, interest in the goal is present at each moment structuring the whole of experience and guiding our activity" (Dreyfus 1992, 275).

Thus, two important differences between factual demands and ontic norms are that the former are (more) definite in their content and apply at a specific time, and the latter are indefinite and apply to Dasein continuously as long as Dasein interprets itself in the relevant way. Indeed, it is precisely *because* the norms are indefinite that they apply to Dasein continuously, and vice versa. Were the norms definite, that is, commutable to precise instructions that specify in addition when and where they should be fulfilled, then one could in principle disburden oneself of the continuous validity of these norms, at least for a while. And conversely, were the norms discontinuous in their validity, then this would mean that they "kick in" in specific forms at specific places and times, that is, that they make up a list of definite demands or instructions. But I will now argue that this is impossible; one cannot reduce the indefinite to the definite in this way and thus disburden oneself of the continuous validity of all the norms.

Suppose that one could (per impossibile)²⁵ reduce being-someone's-friend, for example, to a list of specific things that one

²⁵ This would be impossible, for it would involve knowing the future as well as being able to formulate precisely the content of all the ontic norms and determine in advance how they are to be applied to all relevant future circumstances.

should do on certain specific dates and places (reduce the ontic to the factual). If one could do that, then, given that such a list would be composed of discrete items that describe demands for action that are temporally discontinuous, one would be able to enjoy intermittent respites from the ontic demands that govern one's being-a-friend; these would not apply continuously.

But such a list would be absurd. For supposing one possessed such a list, what would it look like? It would, by definition, include discrete items, discrete pieces of instruction. For example, item "N" on the list might read: "August 28, 2021, 1 p.m.: give your friend a call," and item "N + 1" would read: "September 6, 2021, 6 p.m.: invite your friend over for dinner." But what about the temporal gap between the two instructions? Would one no longer be the other person's friend come August 29? Would one's friendship be terminated at that temporal mark (following the phone call), only to be reignited several days later? This would be absurd, for there is nothing in the fulfillment of that factual demand on August 28 that by itself could explain the temporary termination of the friendship. Indeed, it seems fulfilling a demand of friendship would, everything else being equal, strengthen the normative bond of friendship rather than loosen it. Further, were one to stop being a friend on August 28, then why would one invite the other person who is now no longer one's friend to dinner on September 6? It could be replied that one *does* continue to be the other person's friend (following the phone call), even though one will not have to do anything to realize this friendship until September 6. But, then, in what would being a friend consist? One can insist that it would consist only in the *continuous* and *indefinite* obligation (or meta-demand) to consult the list every now and then to make sure one does not miss the "N + 1" item and perform the next instruction when the time arrives. Not only would this be a bizarre concept of friendship, but, more important, this would in itself show that being-a-friend *cannot* be reduced to a discontinuous list of discrete demands.

A further important element that differentiates the factual and the ontic is that it is precisely because of its relative indeterminateness and open-endedness that Dasein can never be certain whether it is indeed living up to the demands that govern its chosen possibility of Being. Thus, the parent can always ask himself: "Am I a good parent? Am I doing everything I need to do to be a good parent? Am I doing *enough*?" This relatively indefinite nature of the ontic demands is a condition of possibility of such familiar doubts. In contrast, specific factual demands are more definite in their demandingness and typically confront us with more clarity and authority.

Relatedly, and importantly, ontic demands are characterized by a couple of other meta-demands. Thus, one ontic demand that belongs to every possibility of Being (though perhaps with differing degrees of urgency or pertinence) is the meta-demand to figure out *what* the normative demands are that constitute one's possibility of Being. Part of what it is to be a parent, say, just *is* to be concerned with the nature and scope of the demands that constitutes being a parent. Similarly, Being moral inherently requires the continuous watchfulness that demands that one reflect on what one's moral duties are, what is morally right and wrong, and so on. This meta-demand, as ontic, is indefinite in itself, that is, it does not specify in what specific circumstances one should attempt to clarify to oneself the demands that apply to oneself qua being-X, or how. This ontic demand, like all the rest, can, however, materialize in a specific situation. Another meta-demand that can be mentioned here is the demand—which can again become concrete on certain specific occasions—to consider how important this possibility of Being is for oneself and how it measures against others.²⁶

A final crucial difference between the factual and the ontic is that while factual demands can be met and be done away with once and for all (e.g., to read a bedtime story to one's daughter this evening),

²⁶ For this last point, I am indebted to a conversation with Martin Hägglund.

the ontic norms that govern the possibility of being X *can never be met once and for all* as long as one *is* X; in this sense, they cannot be discharged. They always hang above one's head, so to speak. This is because no matter how many factual duties one discharges or commitments one fulfills, one will never discharge the demands themselves that apply to oneself qua X, for first, since the demands are indefinite and cannot be reduced to definite demands, they are in principle impossible to meet by definite actions, and second, to meet the ontic demands in this way, that is, discharge them and thus disburden oneself of them, would entail *ceasing to be* X, given that to be an X means to be governed by the relevant demands.

Let me elaborate on this. In projecting itself onto one or another possibility of Being in terms of which it understands itself, Dasein, Heidegger claims, can only be that possibility in the mode of *to be* (*Zu-Sein*) (e.g., BT, 67, 184). Dasein chooses its possibilities (X, Y), but whatever possibility of being it projects itself onto—and, as thrown, it must project itself in this way—it can never fully *be* X or Y but is always only *on the way to be*: “it is existentially that which, in its potentiality-for-Being, it is *not yet*” (185–186). Put another way, Dasein “*is* what it becomes” (186). This is so because, as existing, Dasein is always “ahead of itself,” where Being ahead of itself means that “in Dasein there is always something *still outstanding*, which, as a potentiality-for-Being for Dasein itself, has not yet become ‘actual’” (279).

Take being a professor. What Heidegger means is that to be a professor is not something that one achieves at a certain point in time once and for all and then just *is*. Put differently, as existing, Dasein, try as it might, can never fully *be* its possibility of Being as completely actual or present; it cannot be, for example, a professor-thing.²⁷ Rather, to be a professor means to live your life so as to

²⁷ To emphasize: Dasein cannot *existentially* be something in the mode of Being of the present-at-hand—this would be a contradiction in terms. It can, of course, be something in the mode of Being of the present-at-hand in the *non-existential* sense. Thus, a given Dasein can be six feet tall. But these are not *existential* possibilities of Being.

maintain your being a professor; it is something one must cultivate and sustain. So Dasein, existentially speaking, is a professor and yet is *not yet* a professor, but not because it is not a professor but a waiter and not because it has not yet been endowed the *status* of professor by the university, but because there is always still more to do to maintain one's being a professor. One's "work" is never done. Dasein's being an X is thus a *task* that Dasein has to work at, where the task is in principle incompletable.²⁸ But what is it that one should do as a professor? One should, for the most part, do the things a professor does. That is, one should act in light of the norms that constitute this for-the-sake-of-which. So, for Heidegger, to be a professor, one must continuously take heed of the relevant ontic norms and act accordingly. But, to repeat, as long as one is an X, *the norms can never be discharged*. It is only once Dasein no longer is a certain possibility of Being that it is no longer *to be* that possibility of Being and is no longer governed by the norms. But what has all of this to do with guilt, ontic or other?

Let's first answer the following question: what does it mean to say that in Dasein there is always something still outstanding? The translators of *Being and Time*, Macquarrie and Robinson, helpfully point out that the German verb Heidegger uses to denote this "outstanding"—*ausstehen*—"is ordinarily used in German to apply to a debt or a bank deposit which, from the point of view of the lender or depositor, has yet to be repaid to him, liquidated, or withdrawn" (BT, 279, footnote 1). Accordingly, I wish to suggest that Dasein, in projecting itself onto a possibility of Being, is guilty—ontically guilty—because, as existing (rather than as Being present-at-hand), it is never fully and once and for all an X but is characterized by "something *still outstanding*," that is, by something which it "owes" to its being an X, so to speak (one's "work" is never done), regardless

²⁸ This does not imply that the task *requires* Dasein to complete it or that Dasein *must* try to complete it and get rid of its *to be*. See also Blattner 2000 for some discussion of incompleatability.

of which possibility of Being it projects itself onto.²⁹ We are now in a position to see why there is reason to call ontic guilt “guilt.” To be *factically* guilty is *not* to discharge a *specific* demand which one ought to abide by in a specific case, a demand that typically concerns others;³⁰ analogously,³¹ to be *ontically* guilty is *not* to be able to discharge, and thus definitely disburden oneself of, the *indefinite*, typically other-regarding demands that continuously govern oneself as X, so that there will be nothing left “outstanding,” to turn all such “ought” into an “is.” *To essentially never discharge the constitutive norms that one is beholden to in one’s possibility of Being is thus to be ontically guilty.*³² Since, however, Dasein is in its very being a thrown projection, it inevitably, essentially, incurs such ontic guilt.³³ Dasein is essentially ontically guilty.³⁴

²⁹ It might seem as if Heidegger is contradicting himself, for, on the one hand, he asserts that “The ‘ahead of itself’ . . . tells us unambiguously that in Dasein there is always something *still outstanding*’ (279), and yet later in the same chapter, he criticizes the idea of the outstanding as applicable to Dasein’s “ahead of itself” (288). To resolve this issue, we need to distinguish between a present-at-hand or ready-to-hand conception of the “outstanding,” which is not applicable to Dasein, and an existential one, which is.

³⁰ As we saw Heidegger puts it, factual guilt is “failure to satisfy some requirement which applies to one’s existent Being with others” (328).

³¹ This, of course, is not a perfect analogy, as should be expected given that we are concerned here with different levels of analysis, namely, the factual on the one hand and the ontic on the other. While in the factual case, one does not discharge a specific demand one should have discharged, in the ontic case, one does not discharge the ontic norms, but in addition, as I argued, one *cannot* (as long as one is an X). This is one central difference between the factual and the ontic.

³² To remove misunderstanding: the claim is not that we try or are required to fulfill the ontic demands completely and fail; rather, as long as we are an X, we exist under norms which continuously apply to us and which we care about meeting—norms that can never be completely discharged.

³³ With the possible exception of the onset of anxiety.

³⁴ Blattner, in a similar vein, interprets “Heidegger’s notion of essential guilt as Dasein’s subjection to norms” (Blattner 2015, 119). Blattner’s reading, though helpful, does not distinguish between the ontic and the ontological as I do here. This prevents, I think, a proper grasp of how the ontological serves as a condition of the ontic, of how ontological guilt makes ontic guilt possible. Specifically, as I will show, ontological guilt runs deeper than Dasein’s subjection to specific norms and concerns Dasein’s caring about itself, which makes its subjection to certain ontic norms possible. Further, Blattner, in my view, does not clearly explain why subjection to norms amounts to *guilt* at all. Indeed, Dasein is subject to certain norms insofar as it projects itself upon a possibility of Being. Projection thus already accounts for subjection. Being-guilty on Blattner’s reading thus appears otiose. Third, though he does recognize Heidegger’s transcendental aim

b. From Factual Guilt to Ontic Guilt: First Step of the Transcendental Argument

Now we can take the argument a step further and see why Dasein's ontic guilt makes factual guilt possible, why it is a necessary condition of possibility of factual guilt: the first-person recognition of being guilty for not abiding by a specific demand one ought to have abided by.

The argument is this. First, since, as ontically guilty, one is continuously governed by certain demands that it is impossible to discharge (as long as Dasein projects itself into the relevant possibility-of-Being), *it is always possible to fail to meet those demands in certain specific ways and acquire factual guilt*. But, second, since to project oneself into a certain possibility of Being is to project oneself understandingly, that is, to see oneself as governed by such demands, it is also possible for one, when a failure occurs, *to see that one has failed to meet these demands in certain specific ways* and thus become aware of one's factual guilt. Dasein's ontic guilt is in this way a necessary condition of possibility of it acquiring factual guilt as well as awareness of this guilt. Put differently, to project oneself onto the possibility of being X means that one will never simply and finally *be* an X (in the mode of Being of the present-at-hand) and rid oneself of the *to-be*. Nevertheless, Dasein, in interpreting itself as X, is thereby committed to *exist* as an X, that is, accepts the normative demands that apply to itself by virtue of interpreting itself as X and tries its best to live up to them. In this way, since Dasein can never finally and once and for all be an X, there will always be a normative gap, so to speak, between Dasein and the discharge of its ontic norms. It is within this gap that specific factual failures can take place and be understood as

in the chapter on the conscience (Blattner 2015, 116), he does not explicitly formulate Heidegger's transcendental argument as I will do here.

taking place, thus burdening Dasein with factual guilt of which it can become aware.

Let us take again the example of being a friend. The argument would then be that given its way of Being, Dasein can never be a friend in the same sense that the table is red, for example; being a friend is not a *state* one acquires at a particular time and continues to possess once and for all (until one dies or stops being a friend). Rather, it is something one continually has to maintain and cultivate; one continually has to rise to the occasion, to meet the expectations, and so on. In other words, being a friend is to exist under the ontic demands of friendship, which become factually determined in various ways and generate specific factual demands in certain specific circumstances. It is precisely because one never is a friend in a present-at-hand way but is always "on the way to be a friend," so to speak, with the horizon always "receding" no matter how much one "progresses"—it is precisely in virtue of this that one is ontically guilty in matters of friendship. But as long as the friend is subject to the relevant ontic demands, she can possibly *fail* to live up to these demands of friendship in certain specific ways, grasp her failure, and consequently acquire consciousness of her factual guilt.

It could be objected that while ontic guilt makes factual guilt possible in this way, it is not *necessary* to be beholden to certain ontic demands in order to be a friend, say, and thus in order to feel guilty as a friend on a specific occasion. Perhaps, so this objection continues, to be a friend is exactly the same as being a non-friend, with the sole exception that when (and only when) the moment arrives, it is only the friend who will behave in the specific proper manner or, conversely, not do so and become, as well as feel, guilty. What makes a person a genuine friend, on this view, is just the disposition to respond appropriately in certain specific situations, much like an automaton of sorts, to offer help, say, when the situations calls for it and to feel guilty for not doing so. Thus, it is not the case that one necessarily must be ontically guilty in order to be factually guilty.

But, it may be asked, how could a “friend” of this kind know that she ought to help, and why would she be motivated to help? More to the point, why would the friend feel guilty for not agreeing to help or not helping? It seems as if she would *necessarily* have to know (at least) both that friends ought to help each other and that she *is* a friend of this other person who is now asking for her help. But this means that she would have to understand herself, at least to some extent (pre-thematically, Heidegger would say), *as* a friend and as such committed to certain broadly defined and indefinite norms that go beyond merely helping in this one specific instance, norms that continually apply and are understood, at least to some extent, as continuously applying to her as long as she is a friend. Indeed, without a grasp of the ontic demands that constitute a possibility of Being, unthematic as it may be, no specific demand could ever manifest itself *as* the demand that it is. But if one is indeed the other person’s friend (and is not hallucinating),³⁵ then one is ontically guilty (as a friend): one is indeed committed to certain norms that cannot be discharged as long as one is a friend. It turns out that one can be and recognize being factually guilty in this way (as a friend on this specific occasion) only if one is ontically guilty.³⁶

It is possible to summarize and formulate the argument that we have so far in the following way:

- (1) Dasein, as existing, that is, as projecting itself onto possibilities of Being, is governed by ontic norms, which it understands.
- (2) Dasein can never discharge the ontic norms that govern it in its projected possibility of Being (to discharge them—which

³⁵ If one were hallucinating, then one would not *recognize* one’s guilt but imagine it.

³⁶ This is not to suggest that one must understand oneself *as* ontically guilty in order to experience guilt. No conception of ontic guilt is required. But whether one understands it or not, the argument is that *being* ontically guilty is necessary for acquiring factual guilt as well as awareness of this guilt.

would be impossible as long as one is still an X—would mean to cease *being* this possibility of Being).

- (3) Which is to say, Dasein is always ontically guilty with respect to these norms.
- (4) It is only because Dasein is ontically guilty that it can fail to do something specific it should have done and also understand this as a specific failure, thus acquiring factual guilt, as well as awareness of this guilt.³⁷
- (5) Thus, ontic guilt, as manifested in the various possibilities of being, is a necessary condition of Dasein's factual guilt, as well as Dasein's awareness of this factual guilt.

An interesting result of this analysis is that from Heidegger's perspective as presented here, the Christian idea of guilt, the idea of being essentially corrupt before God, is revealed, in contrast to Nietzsche's analysis in his *On the Genealogy of Morals*, not as a deepening or "pushing back" (GM II:21) of a genealogically and temporally more original and "superficial" phenomenon of guilt or bad conscience but rather as a specific instance of ontic guilt interpreted theologically as the idea that the human being is essentially deficient; the idea, as Søren Kierkegaard put it, that before God we are always in the wrong. To suffer from Christian guilt is to project oneself onto the Christian possibility of Being, where the relevant normative demand is not merely to be morally good but to be God-like or Christ-like. Since, however, this is *recognized* as an impossibility—it is impossible to discharge one's religious duties and *not* be in the wrong—there arises a belief in an essential fallenness or guilt that belongs to oneself as finite man. To this extent, this Christian interpretation of guilt is exceptionally perceptive; it gives

³⁷ To clarify, even though ontic guilt is characterized by the indefiniteness which I have described, it is not this particular feature that is doing the transcendental "lifting" in each particular case of factual guilt but the fact of being subject continuously to the ontic norms that constitute one's possibility of being.

expression to the idea of ontic guilt that Heidegger considers to belong to Dasein on the ontic level.

On the other hand, the Christian interpretation is deficient from Heidegger's point of view insofar as, first, it tends to regard this ontic guilt as something that man has loaded onto himself as a punishment for his own factual doing (e.g., the consumption of the forbidden fruit, original sin). In other words, it does not clearly distinguish the ontic and the factual. Second, it tends to regard this ontic guilt as playing a *causal role* in bringing about our factual crimes and misdemeanors: it is *because* we are guilty or fallen that we possess the tendency to do evil. Third, it fails to see that this ontic guilt is contingent: it "results" only from one's projection upon the possibility of being a Christian; this guilt does not belong to man as such.

c. From Ontic to Ontological Guilt: Second Step of the Transcendental Argument

I now turn to complete the transcendental argument by showing how *ontological* guilt serves as a necessary condition of ontic guilt. One argument that could be made rather quickly is the following. The more one moves in one's analysis from the factual in the direction of the ontological, the more abstract guilt becomes: a specific demand is conditioned by the ontic demands of a possibility of Being which are in turn conditioned by Being-guilty. The most basic level, that of Being-guilty, is the condition of possibility of ontic guilt, for here we have the fundamental structure that makes possible the guilt that characterizes any specific projection. In other words, Being-guilty, Dasein's ontological guilt, is the condition of ontic guilt because Being-guilty just is an element of the general and basic structure of Dasein, an element that gets filled out with the content of specific ontic projections.

While correct as far as it goes, this answer is far from satisfactory, for first, it does not explain why Being-guilty is Being-guilty; in other words, it does not explain why Dasein's basic character has anything to do with guilt. Second, it does not explain Heidegger's existential definition of ontological guilt as "the null Being-the-basis of a nullity" (BT 353). In what way are Dasein's "nullities" constitutive of Dasein's ontological guilt?

Let me first address the first question. As we saw, to be factually guilty is not to discharge a specific demand, and to be ontically guilty is not to be able to discharge the indefinite demands that constitute a specific possibility of Being. I now wish to assert that for Dasein *to be ontologically guilty is for it not to be able to discharge itself but to be ineluctably beholden or bound to itself*, whatever specific possibility of Being it projects itself onto. We can reach this view of ontological guilt by way of the following propositions.

The first proposition is that Dasein, unlike a present-at-hand thing, is not identical with itself. Rather, as Heidegger puts it in his discussion of Dasein's projection upon possibilities, Dasein, as care, is "beyond itself" . . . as Being towards the potentiality-for-Being which it is itself" (236). This is the ontological structure that makes it possible for us to be always on the way to be what we are (as a professor, parent, etc.). In his explanation of the first nullity (to which I will return in more detail), Heidegger claims that Dasein "*is a nullity of itself*" (330) and a bit earlier that "To this entity [i.e., to Dasein] it [i.e., Dasein] has been delivered over, and as such it can exist solely as the entity which it is; and *as this entity* to which it has been thus delivered over, it *is, in its existing*, the basis of its potentiality-for-Being" (330). In other words, Dasein has been delivered over to itself and is thus not identical with itself or does not coincide with itself but is rather a "nullity of itself," and it is this that is the basis for its potentiality-of-Being. This last claim is important, for Heidegger is basically saying that it is *as this entity* which is not identical with itself that it can be the basis for its potentiality-for-Being, namely, that only as this non-self-identical

entity does it have the existential potential to *be* anything (a teacher, a parent, etc.). This is so since, I want to suggest, it is only this non-self-coincidence that opens the *normative gap* or *space* within which Dasein can exist as something. To see more clearly what I have in mind here, we should turn to the second proposition.

The second proposition is that Dasein, while not identical with itself, is not indifferent to itself. Unlike entities that are present-at-hand to which “their Being is ‘a matter of indifference’” (68), Dasein’s Being is “an *issue* for this entity in its very Being” (67). This means that in the space of non-coincidence between Dasein and itself, there operates a normative force, so to speak: “Dasein exists for the sake of itself” (279). Dasein *cares* about itself.

Dasein, we can say, thus exerts a normative pull on itself but always remains at a distance from itself, never fully coinciding with itself, since in its very Being, it is not self-identical. Indeed, Dasein’s “Being is annihilated when what is still outstanding in its Being has been liquidated. As long as Dasein *is* an entity, it has never reached its ‘wholeness’” (280). Once there is nothing outstanding in its very Being, it no longer *is*, much as once there are no ontic demands that apply to Dasein qua X, it no longer *is* X.

From these claims, it follows that Dasein is ontologically, essentially guilty toward itself or “indebted” to itself. It is precisely because it cares about itself and yet is never self-identical that it is essentially, ontologically, guilty or, what amounts to the same thing, *ineluctably normatively bound to itself*. Dasein is ontologically guilty because it is a normative gap, a gap that yawns between it and itself, a gap Dasein will not—cannot—ever close. Moreover, since it belongs to Dasein as such—it is Dasein’s *ownmost* Being-guilty—and is not “acquired” as a result of factual failure or ontic projection, this guilt is not “an indebtedness which has “arisen” through some deed [*Tat*] done or left undone” (333), and this is in contrast to the metaphysical tradition which, as we saw, precisely makes the argument that it *is* a *Tat* which underlies our essential or ontological guilt.

It is important to see that Dasein's Being-guilty is hereby revealed to be constituted by elements that were introduced already in the analysis in the first division of *Being and Time*, namely, Dasein's ownmost potentiality-for-Being as the ahead of itself which belongs to care, only now we can see these elements under a different light, that is, as constituting Dasein's essential guilt. And indeed, Heidegger says quite clearly that the call of conscience "discloses Dasein's most primordial potentiality-for-Being as Being-guilty [*er erschließt das ursprünglichste Seinkönnen des Daseins als Schuldigsein*]" (334). Dasein's basic constitution is thus disclosed as a kind of guilt.

We can now see why ontological guilt is a necessary condition of ontic guilt. If we were not ontologically guilty, if we did not care about our Being, we could not care about our Being this or that and thus acquire ontic guilt. Looked at from another perspective, focusing on the element of nonidentity, if we were identical with ourselves, then we could not ever become ontically guilty given that the latter involves necessarily lagging behind ourselves (330), never fully and once and for all living up to the demands that constitute us as X or Y.

It is important to clarify that to say that Dasein is normatively bound to itself is not to say that Dasein only cares about itself in egoistic fashion. The claim is *not* that Dasein cares about specific actions because it cares about itself—about its well-being, power, pleasure, and so on. Put differently, the analysis that Heidegger provides of the guilt structure is not *psychological* and is thus not concerned with Dasein's motivations. Dasein doesn't project *in order* to attain some goal or other that concerns itself. Dasein doesn't project in order to reap some reward but insofar as it exists. To hold that it is motivated egoistically in its projections would be to confuse Heidegger's fundamental analytics of Dasein with a specific psychological thesis. Rather, its being normatively bound to itself—its mattering to itself—makes possible its coming to care about X (being a father, say) and thus about specific actions. This self-binding is just a basic existential axiom for Heidegger, the

structure of which gets filled out in various ontic ways. *Why* one projects in this way rather than that or *why* one continues to do so rather than cease and change one's life—these might be given a psychological explanation (though other explanatory factors would also come into play, the psychological not being the only one). For example, I don't care about my daughter's birthday, say, *in order* to be a father, and I don't care about being a father *in order* to achieve some goal that concerns myself. Rather, I care about my daughter's birthday because I *am* her father—caring about her birthday expresses my being a father, we can say—and I care about being her father because that's who I *am*.

Let me now turn to the second question raised above. Why is ontological guilt to be understood as Dasein's null Being-the-basis of a nullity? Asked differently, in what way do these nullities, as Dasein's ontological guilt, condition Dasein's ontic and thus factual guilt? I will start with Dasein's null Being-a-basis.

As we saw in the first section of this chapter, Heidegger explains this nullity as the idea that Dasein can *never* get its basis into its power from the ground up; this "never" constitutes the "not" or "nullity" (330). To show why this idea is an element of Dasein's ontological guilt, it must be shown how this nullity is a necessary condition for Dasein's ontic and factual guilt. Since, however, it is impossible to get a handle on what exactly getting one's basis into one's power would mean at the ontological level, given that it abstracts from *all* content, factual or ontic, I think it would be helpful to ask what it would mean for Dasein to be in control over its *ontic thrownness*. Seeing what this amounts to can enable us to see what it means for Dasein to be guilty as such, that is, at the ontological level.

Dasein, as Being-in-the-world, is thrown into a world not of its own making, a central characteristic of which is that it makes available for Dasein certain possibilities of Being rather than others upon which it can project itself. So what would it be for Dasein to "come back behind its thrownness" or to have "power over one's

ownmost Being from the ground up" (330)? It would mean *at least* the following. It would mean, first, to be in a position to decide *whether* there is available for Dasein a certain possibility-of-Being at all, that is, whether such a for-the-sake-of-which is part of one's world and is thus, as William James put it, a live option. Second, it would involve the ability to decide exactly *what* ontic norms are constitutive of this possibility-of-Being and how specifically they are to be applied to any and every specific situation. It would therefore be to be able to determine beforehand and exactly what should be done in any given kind of situation. This would be to eliminate all normative indefiniteness and would amount to making "an inventory of . . . [Dasein] . . . as something-to-hand and list the contents of its Being" (185). It would also eliminate one central norm that, as we saw, constitutes each possibility of Being, namely, the meta-demand to determine the normative content of one's possibility of Being; since one would fully determine oneself, there would be no room for this ongoing meta-search.

Why, then, would having power over oneself from the ground up eliminate one's Being-guilty and consequently the possibility of *all* guilt, ontic as well as factual? To have power over one's thrownness, I would like to suggest, would mean to be able to create for oneself a possibility of Being on the spot and be able to determine its ontic norms by oneself and thus *never be guilty*. This is because were one to be able to determine one's "there" by oneself, one could always make it so that there is *nothing* outstanding with respect to one's ontic norms and consequently with respect to any factual demand. One could make it so that every time, one is exactly what one ought to be as an X, *without any remainder*. One would then not lag behind one's possibilities (330) but coincide with them perfectly. The gap between the "ought" and the "is" would close, and one would then just be what one is and attain, or almost attain, the mode of Being of the self-identical present-at-hand. For example, possessed with such a power over oneself, one could decide in a given moment that one is a "shmarent" (in contrast to a parent) and that to be

a shmarent *just is* to slouch at present before the computer and do whatever one wishes to do and proceed to do that. More radically, it would mean to have the power to create for oneself the possibility of being “God,” where one is at every moment exactly what one should be.³⁸ It would then be impossible to be ontically and factically guilty as a “shmarent” or as “God.”³⁹

But Dasein, of course, is not like that—we are never *causa sui*. We are not the kind of Being that can fully determine itself by determining its ontic possibilities of Being.⁴⁰ Rather, we are constantly behind ourselves, constantly lagging behind our possibilities of Being the content of which is not determined in this way by us. We thus never fully, completely coincide with ourselves but are ontologically guilty—essentially *normatively bound*. And now we can see the deeper reason for why being *causa sui* in this way would eliminate the possibility of ontic and factual guilt, namely, it would eliminate Dasein’s ontological Being-guilty, which, as I have argued, is the necessary condition of both kinds of guilt. For consider: were Dasein in control of its own thrown Being from the ground up in the manner I have described, it would lose its normative nature. And yet Dasein is essentially, as Being-guilty, normatively bound to itself. Further, Dasein would lose the gap that separates it from itself and collapse into itself. And yet Dasein, as Being-guilty, is

³⁸ Compare to the Descartes epigram for this book.

³⁹ Heidegger’s argument here is reminiscent of Wittgenstein’s private language argument where the point is made that were we to decide the meaning of each word by ourselves, there would be no sense to talk about right or wrong.

⁴⁰ While it is true that we are not *causa sui* in this sense, an interesting question is why. Why can’t we decide that we are “God” (in the sense explained in the text) and live our lives accordingly? From a Heideggerian perspective, the answer must lie in the kind of being that we are: we are Dasein, and Dasein is a null basis, not a *causa sui*. This is correct as far as it goes but is perhaps not informative enough. A more illuminating response, perhaps, would involve thinking in concrete terms about such a case. Specifically, if we could decide that we are “God” and create ourselves as one as *causa sui*, we would still have to think about what to do in the next instance of our godly existence, and in order to answer this practical question, we would inevitably have to refer to certain norms that we take to have some authority over us as the divine being that we are and try to figure out how they apply factically to the situation before us. In other words, it seems, we would have to acknowledge our ontic guilt.

nonidentical to itself. It thus follows that Being a null basis is constitutive of Being-guilty. We thus reach an important conclusion of our analysis of Heidegger's discussion of guilt: *Dasein's not being causa sui is a necessary condition of the possibility of guilt überhaupt.*⁴¹

So far, I have focused on one of the nullities that constitute ontological guilt—"the null Being-the-basis." But Heidegger, as we saw, thinks that another nullity is at issue, namely, that Dasein is also a basis of a nullity. How does this figure into Dasein's ontological guilt? Further, as we saw at the beginning of our analysis, according to Heidegger, the formal analysis of the everyday understanding of guilt concerns two *interrelated* "lacks": "*Being-the basis* for a lack of something in the Dasein of an Other, and in such a manner that this very Being-the-basis determines itself as 'lacking in some way' in terms of that for which it is the basis" (328). Since the ontological analysis of guilt is supposed to grasp the everyday conception in a more primordial, even further formalized manner, we should expect the two nullities that constitute Dasein's ontological Being-guilt to manifest some sort of interconnectedness as well.

Heidegger explains that "Not only is the projection, as one that has been thrown, determined by the nullity of Being-a-basis [first nullity]; *as projection* it is itself essentially null [second nullity]" (331). The nullity has to do with the fact that "in having a potentiality-for-Being [Dasein] always stands in one possibility or another: it constantly is *not* other possibilities" (331). And Heidegger continues to clarify that this has to do with Dasein's

⁴¹ Lee Braver explains this primordial guilt as follows: "We did not decide to be born, or where or when or as what, *nor did we enact our own creation*" (Braver 2014, 86; emphasis added). In the italicized part, Braver seems to express the same view I articulate here, where Being-guilty precisely means not being *causa sui*. Braver, however, continues as follows: "we are, before our first breath, beneficiaries of people and events that we had no part in even though we owe our very existence to them. This inescapable indebtedness is the 'not' or nullity that lies at the very basis of our being anything at all" (86). This exegesis is problematic because it turns ontological guilt into factual guilt, where Dasein is conceived as indebted to specific actions of specific others. As we saw, Heidegger stresses how the primordial sense of Being-guilty makes no reference to the specific acts of others and does not involve the factual balancing of debts (BT 328).

"Being-free for its existentiell possibilities" (331). One way to read this is to view the second nullity as complementing the first: since Dasein as such is *not* determined by a specific possibility or a set of possibilities and *cannot* determine its possibilities by itself for itself (first nullity), it has to and can project itself into this or that possibility. Dasein—to put it in Bernard Williams's term (1993b)—has no necessary identity and cannot create one for itself as *causa sui* but is nevertheless free to be this or that in its projection.

I therefore suggest that the two nullities are connected in the following way: it is precisely because one has no control over one's Dasein from the ground up (the first nullity) that Dasein is "null *in advance* of any of the things which it can project" (BT 331) and is thus not ontologically (and consequently ontically) characterized by any specific possibility of Being. To put it differently, it is because one is a null basis—never has power over one's Being from the ground up—that one is a basis of a nullity, that is, is *not* essentially determined by a specific possibility of Being but is free to project oneself into this or that possibility. Had one such power, one could determine oneself as X or Y in advance in the manner I have described and would thus not have to project oneself as X or Y in one's thrown existence, with all the indeterminacy and open-endedness that such a projection entails.⁴²

There is, however, another (though compatible) way to understand this second nullity and its connection to the first. I argued that Being-guilty involves the idea that Dasein can never become identical with itself but always remains within the normative gap that separates itself from itself. Now, the idea of being a basis of a nullity—the second nullity—could be read as the idea that no

⁴² Notice that this indeterminateness of Dasein as such is also the condition of possibility of Dasein's Being inauthentic. As Heidegger puts it: "in the structure of thrownness, as in that of projection, there lies essentially a nullity. This nullity is the basis for the possibility of *inauthentic* Dasein in its falling" (331). In other words, the fact that Dasein is not fixed but can be this or that makes inauthenticity—as well as authenticity, it should be added—possible.

matter *which* possibility of Being Dasein projects itself upon, it will *not* become self-identical and thus rid itself of its ontological guilt. No matter what Dasein does factually, it will never rid itself of its ontic guilt (qua X). Similarly, no matter what possibility of Being Dasein projects itself upon, it will never rid itself of its ontological guilt. As Heidegger explains, this second nullity relates to freedom, that is, “the choice of *one* possibility” which involves “tolerating one’s not having chosen the others and one’s not being able to choose them” (331). But this “tolerating” “*not* [choosing] other possibilities” cannot be the factual idea that I as a specific Dasein feel some kind of regret for choosing to *be* this rather than that (a programmer rather than a freedom fighter, say), for this regret is not something that belongs to Dasein as such at the ontological level, but Being-guilty and the nullities that constitute it do. Consequently, this “tolerating,” I suggest, has to do with the idea that Dasein is constituted ontologically in such a way that *whatever* it chooses existentially, it remains ontologically guilty and *has to tolerate that*. It is the Kierkegaardian idea that one is going to regret whatever one does⁴³ but grasped *ontologically*. And here, too, with respect to Dasein’s ontic projections themselves, just as in the case of the norms that make up each projection, there is indeterminateness: the norms that constitute a possibility-of-Being, as we saw, are indefinite, but the norms that can guide one’s projection (what should I become? who should I be?) are also indeterminate, perhaps even more radically so.⁴⁴

If my analysis here is correct, then it follows that in Heidegger’s analysis of Being-guilty, a *reversal of reasoning* takes place: while for the metaphysical view, factual guilt presupposes ontological guilt,

⁴³ Kierkegaard 1992, 54–55.

⁴⁴ One problem with Heidegger’s talk of Dasein’s freedom to project itself upon its possibilities of Being is that in many cases, the possibility of being onto which Dasein projects itself is not the result of free choice. And yet Dasein can find itself burdened with the ontic guilt of, say, being Jewish, being a man, and so on. So there is an element of voluntarism in Heidegger’s talk about choice, which stands in some tension with the idea of thrownness.

which in turn presupposes our being *causa sui*, for Heidegger, it is exactly our *not* being *causa sui* that constitutes our Being-guilty, which in turn makes possible our factual guilt.⁴⁵ Importantly, Heidegger agrees with Nietzsche that we are not *causa sui*, but he still holds that we are ontologically guilty and thus can become factually guilty. Heidegger, like Nietzsche, disagrees with the metaphysical tradition that we exist on some noumenal level, but unlike Nietzsche, he does not think that this immediately entails that we cannot be genuinely and justifiably guilty. If the reversal of reasoning I have here presented is successful, Heidegger can claim to have avoided a serious objection against the notion of factual guilt: guilt is still potentially attributable to us even if we are not *causa sui*.

Before I proceed, I wish to respond to an objection according to which my interpretation of Being-guilty runs counter to the text, for Heidegger specifically explains that “The idea of guilt must not only be raised above the domain of that concern in which we reckon things up, but it must also be detached from any relationship to any ‘ought’ or law [*Sollen und Gesetz*] such that by failing to comply with it one loads himself with guilt” (BT 328; translation slightly modified). And yet the interpretation suggested wishes to construe guilt in terms of demands and norms, which sounds suspiciously close to the concepts Heidegger wishes to do away with. The response to this objection is twofold. First, the ontological Being-guilty in itself, on the interpretation I have offered, makes no reference to any *content* that demands compliance; the normative space marked by

⁴⁵ In his paper on the subject, Bernd Irlenborn claims that with Heidegger’s existential analysis of guilt, “Heidegger reverses the usual relation that holds with respect to the concept of guilt and can therefore identify Being-guilty as a condition of possibility for indebtedness. In contrast, the familiar conception of guilt sees guilt as grounded only in a specific debt, that is, in an offense against a certain norm” (Irlenborn 2004, 196; my translation). I agree with Irlenborn that there is a reversal here, but the specific reversal he mentions in this quotation is not new with Heidegger; as we saw, it can be found already in the metaphysical tradition, where factual guilt is grounded in a more primordial or ontological guilt. The more radical reversal Heidegger conducts is the one I have identified here.

ontological guilt gets filled out, so to speak, when Dasein projects itself ontically onto this or that possibility-of-Being. Second, the “ought” that Heidegger has in mind in the cited text is related to that which “is still necessarily defined as a *lack*—when *something* which ought to be and which can be is missing” (328; second emphasis added). In other words, the “ought” is a *factual* ought, one that arises in specific situations and *can* be fulfilled so that failing to live up to it can result in factual guilt. In contrast, Being-guilty as well as ontic guilt make no reference to any particular *something* that ought to be, can be, and is missing or lacking. Thus, in the case of ontic guilt, there is no something that can be done that will eliminate one’s guilt. Ontic guilt is continuous as long as one is a specific possibility of Being. Similarly, in the case of ontological guilt, there is no something that is missing and can be done or achieved that will close the gap and finally unite Dasein with itself.⁴⁶

IV. The Phenomenon of the Call of Conscience

I reconstructed Heidegger’s argument for the claim that our ownmost Being-guilty is a condition of possibility for factual guilt. In addition, I explained why this Being-guilty can properly be seen as belonging to Dasein as such, namely, to its null Being the basis of a nullity. Nevertheless, we can still wonder, *are* we indeed Being-guilty? Is this the correct way to think about our Dasein (if we are indeed Dasein at all)?

It is possible for us to quickly answer this question in the affirmative: given that we *do* experience guilt,⁴⁷ it follows from the transcendental argument I have reconstructed (assuming its

⁴⁶ Suicide does not close the gap but eliminates it. It does not unite Dasein with itself but annihilates Dasein.

⁴⁷ Whether or not this experience is fully justified is a different question I am now putting aside.

soundness) that we are indeed Being-guilty. Heidegger, however, does not explicitly provide us with such an argument and seems to rely (at least to some extent) in his discussion on a phenomenological analysis of the call of conscience, one that is supposed to reveal phenomenologically to us our Being-guilty. As he claims: “the call gives us *this Being-guilty* as something which at bottom we are to understand” (BT 332). Heidegger’s phenomenological findings—to remind the reader—are that the call of conscience “comes *from me* and yet *from beyond me and over me*” (320), that in conscience Dasein calls its *own Self* (317) from the “they-self,” and that this calling is a “mode of discourse” (316). Moreover, in the call, Dasein, as caller, “*finds itself in the very depth of its uncanniness*,” and “[u]ncanniness reveals itself authentically in the basic state-of-mind of anxiety” (321), where Dasein is “anxious about its potentiality-for-Being” (322). Finally, the call, though it “does not report events” and “calls without uttering anything” (322), gives Dasein something to understand, namely, that Dasein is “Guilty” (e.g., 325)—it calls Dasein *to* its ownmost Being-guilty which “remains closed off from the they-self” (334). These various claims, he seems to hold, are the “phenomenal findings” (320); this is what we “grasp phenomenally” (325), and “all those phenomena which we have set forth in characterizing the caller and its calling speak for it [i.e., for the claim that Dasein, in the very depths of its uncanniness, is the caller of the call]” (321). I believe that we are now in a position to better make sense of these claims. Let me quickly go over these elements.

The call of conscience “comes *from me* and yet *from beyond me and over me*.” This is to be expected if Being-guilty is to be conceived of as the normative gap that yawns between Dasein and itself. The gap as such, in abstraction from any specific ontic possibility of Being, lacks all content, which is why the call of conscience says nothing. But it gives Dasein something to understand, namely, its Being-guilty; it reveals to Dasein its essential guilt; it reveals the gap as gap. In the call of conscience, this gap reverberates—is made to

sound. Furthermore, the call reveals Dasein in its anxiety. Though I cannot enter here into a detailed analysis of anxiety, if we take anxiety as Dasein's revelatory awareness of the empty normativity of Dasein as such, an emptiness that is disturbing, for Dasein craves normativity—Dasein, after all, is *care*—then we can see why the call of conscience is characterized by anxiety. At the ontological level of Being-guilty, Dasein is face to face with its mere boundedness to itself in abstraction from any specific norms. This detachment from any specific normative commitments alongside the remaining need for normative guidance is productive of anxiety.

And yet, despite Heidegger's assertions that this analysis is grounded phenomenologically, it is not clear what actual phenomenon manifests the elements that Heidegger describes. Heidegger seems to be aware of this problem and returns to it in various forms again and again. Thus, he asks, "But what kind of experience speaks for this primordial Being-guilty which belongs to Dasein?" (332), and later, "can we claim sufficient evidential weight for the way we have exhibited this [potentiality-for-Being]?" (335). Heidegger seems to address this concern in section 59, which is titled "The Existential Interpretation of the Conscience, and the Way Conscience Is Ordinarily Interpreted." Here Heidegger discusses the various differences that separate our everyday experience and interpretation of conscience and the existential call of conscience. But this in itself still gives no reason to think that the call that Heidegger describes is something that we ever come across; typically, guilt, moral or otherwise, is not characterized by reticence and anxiety at all.

In response, Heidegger can argue that given that typically we are lost in the "they" and so interpret the conscience in an everyday, ontologically superficial manner, it should come as no surprise that we don't come across the call of conscience as existentially interpreted. Even more sharply still, Heidegger, it seems, wants to make the claim that it is precisely *because* the call calls us back to ourselves that we cannot hear it in a way that goes beyond the

everyday interpretation, for here we have a “direct conflict” (335) between the two, as if the everyday interpretation, adopted by the “they-self,” actively resists the proper hearing of the call so as not to lose its grip on Dasein and let it become authentic (335). This is hardly convincing and strikes this reader as a radical attempt at a hermeneutic of suspicion of sorts that begs the question; it is hardly satisfactory as an argument for the existence of X to claim that we don’t see it and can’t find evidence for it, because experiencing it would be too harmful to us in some way and that therefore some active force in us makes sure that it remains hidden from sight.

An alternative way to address the problem of phenomenological evidence, or lack thereof, would be to maintain not that the everyday interpretation of the call of conscience *completely* occludes the deeper existential phenomenon but that it merely obstructs access to this deeper layer as a phenomenological muffler of sorts. According to this way of thinking, we are essentially dealing with *one* phenomenon here, namely, that of the call of conscience as existentially understood, with the everyday interpretation limiting our interpretation of this phenomenon. Indeed, Heidegger claims, though rather tentatively, that the factual feeling of guilt or “factual indebtedness were only the occasion for the factual calling of conscience” (337), that, in other words, the latter expresses itself in the former, though it typically cannot be clearly heard. As Heidegger puts it, “In the common-sense interpretation, one may suppose that one is sticking to the ‘facts’; but in the end, by its very common sense, this interpretation has restricted the call’s disclosive range” (340; see also 347). Everydayness prevents us from expanding our disclosive range and hearing the call down to its depth, grasping it ontically and—even further broadening our hearing—ontologically.

It thus seems that the analysis of factual guilt should be thought of as analogous to Heidegger’s analysis of the breakdown of the ready-to-hand: when something goes wrong and things cease to work, we get a glimpse of what has been the case all along, we get

a glimpse of the kind of being of the ready-to-hand, a glimpse that we can pursue and thus deepen our interpretation. Similarly, when we fail to meet a specific demand and experience factual guilt, this shows that something has gone wrong but also, at the same time, arguably provides a glimpse of what has been the case all along, namely, our Being-guilty. We can then, if we are primed to hear the call properly or are engaged in fundamental ontology, hold on to the end of the string and follow it down the path to a proper hearing of the call and to our ownmost Being-guilty.

But the two cases are not perfectly analogous, for in the case of the ready-to-hand, we can return, post-breakdown, to attend to how equipment phenomenologically presents itself and confirm our analysis regarding its unobtrusive nature, its being part of a network of assignments that give it meaning, and so on. On the other hand, in the case of the call of conscience, the phenomenological evidence to support the ontological interpretation that Heidegger presents us with would still be missing. Is there an experience where the reticence, the anxiety, and the "Guilty!" could be phenomenologically confirmed?

One possible idea is that sometimes—perhaps under certain conditions such as when the crime is great and the punishment not clearly forthcoming—the experience of guilt could be characterized by a feeling of horrific nothingness, a realization that Dasein as such is not governed by *any norm*. Possible fictional examples of this are Judah in Woody Allen's *Crimes and Misdemeanors* and Klytämnestra in Richard Strauss's *Elektra*.⁴⁸ But what makes this experience one of *anxiety* is precisely that Dasein is in its essence *normative* and thus seeks to be governed by some norm or other. Anxiety is the craving for normativity in the face of the nothingness

⁴⁸ In Strauss and Hugo von Hofmannsthal's *Elektra*, Klytämnestra gives voice to this silent nothing that horrifies her thus: "kriecht zwischen Tag und Nacht wenn ich mit offenen Augen lieg' ein Etwas hin über mich. Es ist kein Wort, es ist kein Schmerz, es drückt mich nicht, es würgt mich nicht, *nichts* ist es . . . und dennoch, es ist so fürchterlich, dass meine Seele sich wünscht, erhängt zu sein."

of the world. Anxiety is possible *only* because Dasein is normative. Otherwise, there would be nothing essentially disturbing in the realization that “there is nothing that binds me” (Kierkegaard 1992, 51); there would be no anxiety in the anxiety. Anxiety is the existential vertigo a normative being experiences in the face of a normative vacuum.⁴⁹ It is unpleasant much like an attempt to breathe in space. Dasein is a normative gap, and in anxiety the gap is emptied out, and Dasein experiences its existential *horror vacui*. The typical reaction is to flee from anxiety and throw oneself into the consoling embrace of the “they.” In the case of guilt for a crime, the response would be to turn oneself in and let the law fill out the suffocating, empty normative space. Fyodor Dostoevsky’s Raskolnikov comes to mind.

If these observations are on the right track, then they would seem to suggest an alternative way to understand the intriguing phenomenon of the need for punishment that the culprit might experience, a need that might compel the culprit to seek it out. While, as we saw in chapter 4 on Rée, a psychological explanation of this phenomenon can appeal to psychological associations formed in the process of education, an existential, Heideggerian account can invoke the unbearable normative emptiness that an unpunished crime can give rise to.

It does not seem, however, that the result of this analysis should be satisfying. For it now turns out, if the phenomena I have adduced here can be regarded as evidence pertinent to Heidegger’s description of the call of conscience, that the person most acquainted with the ontological nature of the call is not exactly *authentic* Dasein but *criminal* Dasein. It is the person who, having committed something morally atrocious, manages to get over or get beyond the factual dimension of guilt, concerned as it is with specific acts and specific debts, reaches the heart or the bottom of the phenomenon

⁴⁹ Literal vertigo as a factual experience is similar: I value my life but realize there is nothing that binds me—I am free to throw myself off the cliff and end it.

and is terrified by what he or she finds. The result of properly hearing the call of conscience would then be not authentic resolve but handing oneself over to the police (Raskolnikov), going mad (Klytämnestra), or returning to one's normal life (Judah).

V. Further Conditions of Guilt and Wanting to Have a Conscience

I have so far argued that Being-guilty is a *necessary* condition for factual guilt. But why is it not also sufficient? What is still missing? In this section, I will look at several deficiencies of Heidegger's view that have to do with his lack of philosophical engagement or concern with factual guilt, its nature, and its conditions.

Heidegger is mostly silent, first, on the *feeling* of guilt, moral or other. He is silent about its specific qualitative character, namely, the unpleasantness that it involves—no explanation of this is given in Heidegger's *Being and Time* or elsewhere. Neither does he say anything about the psychological process that instills in agents the capacity to feel factual guilt in the first place.

A related issue is the following. We saw that in the metaphysical tradition, ontological guilt not only makes factual guilt possible by providing the conditions of the *imputability* of actions in the substantive sense, but it also *explains* why we act in the ways that we do, that is, morally or immorally. In Heidegger, both dimensions are lacking. With respect to the first dimension, while Heidegger does provide us with an exploration and clarification of the ontological and ontic conditions of possibility of factual guilt, he does not analyze the conditions under which a specific action can be imputed in a substantive sense to an agent. In other words, he does not focus on the conditions of factual *responsibility*, namely, on what has to be the case so that an individual agent can be rightly seen as responsible for a specific action performed under specific circumstances and thus be rightly considered by himself or others as guilty for

the action (in case the action is wrong in some sense). As we shall see, however, Heidegger *does* address the question of existentiell or ontic responsibility.

With respect to the second dimension found in the metaphysical tradition, Heidegger elucidates the ontological and ontic conditions of possibility of Being-guilty in the factual sense, but he does not concern himself with questions of explanation. Stated differently, Being-guilty and ontic guilt do not explain *why* we tend to act in ways that give rise to factual guilt. And from Heidegger's point of view, at least, this is properly so, for insofar as these questions—psychological questions—are taken to be basic, then they are based on a fundamental misunderstanding of the kind of Being that we have, given that psychology attempts to understand Dasein as if its kind of being were that of the present-at-hand, namely, as an arena for the acting on each other of various forces. From Heidegger's perspective, psychology—as it is also found in Rée and Nietzsche—amounts at best to a derivative, and thus a deficient, perspective. But surely a philosophical position that as a matter of principle denies the possibility of explanation or cannot clearly make room for explanation when it comes to the human being and its behavior is deeply flawed (see Golob 2014, 211, for a similar complaint).

Let us now turn to examine Heidegger's idea of “wanting to have a conscience.” Factual guilt presupposes *responsibility*: to consider oneself or be considered by others as justifiably guilty for a commission or an omission necessarily presupposes justifiably considering oneself responsible for such a commission or omission. Heidegger, as I claimed, seems to have little to say about factual responsibility, its nature, or its conditions. But he *does* seem to think that a condition of possibility for being a responsible X or Y (a responsible father or nurse), that is, for being responsible *existentiell*y, concerns the phenomenon of wanting to have a conscience, which he describes as a response Dasein can have to the call of conscience, the call that confronts one with one's ownmost Being-guilty. As Heidegger puts it:

Wanting to have a conscience is rather the most primordial existentiell presupposition for the possibility of factically coming to be guilty/to owe something [Schuldigwerdens]. In understanding the call, Dasein lets its ownmost Self take action in itself in terms of that potentiality-for-Being which it has chosen. Only so can it be answerable/responsible [verantwortlich]. (BT 334)

This is a complex passage where the ideas of Being-guilty, ontic *responsibility*, and factual guilt are explained in their interrelationships. The idea seems to be that one can only become responsible/answerable “in terms of that potentiality-for-Being which [one] has chosen” (334), that is, existentially or ontically (as a father, say), and thus potentially genuinely factically guilty for a specific action, when one wants to have a conscience, that is, when one both listens to the call of conscience and projects oneself onto this ownmost Being-guilty, that is, when one acquires *resoluteness* (*Entschlossenheit*), a central element of authenticity or ownedness (e.g., 343). I cannot unpack these complex notions here, but it thus seems that besides Being-guilty and ontic guilt, there is an *additional* necessary condition Dasein must meet in order to be genuinely guilty for a specific misdeed: wanting to have a conscience.

But this is problematic. If wanting to have a conscience is an element of authenticity (anticipation of death being the other central element; e.g., 307), and if authenticity is either a rare or a precarious achievement since “Dasein is already in irresoluteness, and soon, perhaps, will be in it again” (345), then it is not clear how third-person or first-person ascriptions of factual guilt can “stick” in the manner in which we might want or expect them to. In other words, this approach threatens to get people off the hook, for, first, it will be quite impossible, at least from the “outside” but perhaps also from the “inside,” to determine whether anyone is genuinely factically guilty given that authenticity does not manifest itself in clear and distinct behavior. In addition, second, whether someone is or is not factically guilty will depend not on any action or decision for

which one can be held responsible but on the arrival of the call of conscience and the proper response to the call. In other words, for Heidegger, it seems, one isn't responsible for becoming existentially responsible, and thus one cannot be *held* responsible for one's failure to become so responsible. Consequently, one cannot be held responsible for one's factual misdeeds. The miscreant cannot be justifiably held responsible for her misdeed if responsibility did not *happen* to her, so to speak. Consequently, she cannot fully justifiably be seen as factually *guilty* by herself or others.

Another suggestion for making sense of Heidegger's idea of wanting to have a conscience is the following. It can be argued that "wanting to have a conscience" involves recognizing the validity and authority of certain principles over oneself, in line with (some of the) compatibilist criteria of reflexive control (Wallace 1994). As we saw, however, projecting oneself onto a possibility of Being already includes understanding and caring about the norms or values that govern oneself and thus one's relations to others, so why is there needed a further moment—wanting to have a conscience—of recognizing the validity and authority of these values or norms? Doesn't everyday Dasein recognize these even "before" it comes to properly hear the call of conscience?

In his discussion of the conscience in *Being and Time*, Steven Crowell addresses Robert Pippin's objection (Pippin 1997) according to which it is not clear where in Heidegger there is room for Dasein acting not merely in accordance with but *in light of* norms (Crowell 2007, 46–47). Crowell then attempts to show that in Dasein's response to the call of conscience, there is a moment where Dasein takes over being a ground existentially such that "factic grounds become subject to a choice for which I am accountable; they are thereby taken up into the normative space of reasons" (57). But this way of reading wanting to have a conscience, I think, has the following troubling implication: if it is the case that before the proper response to the call, Dasein could only act in accordance with but not in light of norms, then this would imply that everyday

Dasein, in its inauthentic mode, is, as Crowell puts it, a “mindless” (46) zombie that only reacts to its surroundings in various conformist ways that can be described from the outside as moral or proper but does not see that in certain situations, it *should* or *ought to* behave in some way. This would have the further implication that inauthentic Dasein could not be held responsible for its actions: zombies are off the hook.⁵⁰

There is, however, no reason to push ourselves into this corner in this way. Dasein surely can act in light of norms independently of its responding properly to the call of conscience; it has a kind of *sight*—circumspection (*Umsicht*)—involved, for example, in the carpenter’s ability to see that a certain tool is not going to be *proper* or *good* for a certain task (“Too heavy!”). This is why the activities that Dasein performs in the world are not blind: “But when we deal with them [things] by using them and manipulating them, this activity is not a blind one; it has its own kind of sight, by which our manipulation is guided” (BT 98). In other words, it belongs to Dasein’s constitution that its activities are not mindless but are *guided* by a seeing—*normatively* guided. It therefore does not seem that Heidegger’s discussion of conscience in general or the idea of wanting to have a conscience in particular is needed to fill this alleged gap in Dasein’s existential analysis.⁵¹

⁵⁰ There are many other issues that arise here. For instance, Crowell claims that Being-guilty “identifies the ontological conditions whereby one’s (factic) grounds become my normative (reasons) and thus explains how Dasein can act not only in accord with norms but also in light of them” (Crowell 2007, 58). It seems the view here is that in the call of conscience, which Crowell does not really distinguish from the state of mind of anxiety, I take a step back from my nature (Crowell talks about the ground being nature, e.g., in Crowell 2007, 58) and then can take it *as* a reason, which in Kantian-speak basically means incorporating it into my maxim. Not only is this to Kantianize Heidegger, but it also is to imply that only authentic Dasein acts in light of reasons. Inauthentic Dasein, lost in the “they” (*das Man*), on the other hand, is like an animal: it just reacts causally to what is going on. This, I think, is false phenomenologically.

⁵¹ I believe Pippin’s challenge to Heidegger has to do, ultimately, not merely or not exactly with Dasein’s ability to act in light of norms as such but with Heidegger’s inability to make sense of how norms come to be in the first place, of how they come to be authoritative over Dasein, and of how they come to be supplanted by different norms. In Heidegger’s view, we just find ourselves thrown into a world already governed by norms, and we learn to orient our Being-in-the-world in light of them or choose to subscribe

It is possible, despite the difficulties noted, to claim that Heidegger's analysis in the passage on page 334 is quite appropriate and that this is how it should be, namely, that genuine factual guilt can only be possible where there is wanting to have a conscience, only where there is authenticity or ownedness. It is only when Dasein has projected itself upon its ownmost Being-guilty that it becomes truly existentially responsible or answerable and can thus genuinely and justifiably be guilty for its factual omissions or commissions. In contrast, as long as Dasein does not want to have a conscience, even though it might break the law or transgress the dictates of morality or even come to *feel* guilty about it, none of this would make Dasein genuinely responsible and thus genuinely guilty. This is so because—this suggestion continues—as long as Dasein is inauthentic, even though it is familiar with the norms, recognizes their validity upon itself, and can also act in light of them, it is nevertheless not genuinely responsible because it has been disburdened (BT 165) of its responsibility by the “they” and “*dispersed* into the ‘they’” (167); its “possibilities of Being” have been taken away by the Others, “to dispose of as they please” (164). More specifically, it is the “they” which “prescribes the kind of Being of everydayness” (164). It commands the content of the ontic norms that constitute Dasein's various possibilities of Being as well as how they are to be applied to particular situations. In this way, the “they” “deprives the particular Dasein of its responsibility/answerability [*nimmt es dem jeweiligen Dasein die Verantwortlichkeit ab*]” (165). And, most strikingly, “In Dasein's everydayness the agency through

to them (when we adopt a certain role), but the norms themselves arise not as a result of a Hegelian dialectic but, so we learn more clearly in Heidegger's later work, in an *event* of truth, in a happening of *aletheia* or unconcealment. Similarly, norms change or disappear not as a result of the rational progression of Spirit but as a result of a world withdrawing into concealment. In this sense, Pippin is not completely wrong, I think, to claim that in Heidegger, norms appear/collapse in a “comprehensive” and “discontinuous” manner (Pippin 1997, 390). Their collapse, however, is not the result of anxiety, as Pippin seems to suggest (see also 392); anxiety does not lead to the collapse of the *world* and its norms but only to its becoming nothing for *me*, and only episodically, not terminally.

which most things come about is one of which we must say that ‘it was no one’ [*In der Alltäglichkeit des Daseins wird das meiste durch das, von dem wir sagen müssen, keiner war es*]” (165).

In contrast, to become resolute, to project oneself upon one’s ownmost Being-guilty, that is, to properly hear and respond to the call of conscience, means to take one’s responsibility back from the “they”; it is to want to have one’s *own* conscience. The fallen Dasein of everydayness, lost in the “they,” has a conscience, too—it can, after all, suffer factically from bad conscience or enjoy its good conscience, but its conscience is not its own insofar as it tends to interpret its ontic demands and the specific factual demands that these issue in line with the “they.” To become authentic or one’s own, to appropriate oneself, is to want to win oneself back from one’s lostness in the “they.” This entails the will to appropriate one’s conscience. More concretely, what this involves is recognition of the “indefiniteness” (345) of the ontic ought and the corresponding readiness to interpret it one’s own way and discover “first what is factically possible” (346). It thus involves *taking responsibility for the ontic norms themselves—for how they are to be interpreted and applied factically*.⁵² It involves, in other words, seeing that the norms are up for grabs—by *me*.

As interpreted here, the idea of wanting to have a conscience as a mark of authenticity is strikingly similar to the idea represented by the figure of the sovereign individual that we encountered in chapter 5 on Nietzsche. In both cases, what is at issue is the individuation of a unique individual who attains his or her own conscience as a result of tearing himself or herself away from some form of a

⁵² This is close to Crowell’s analysis as I understand it, but with the following important difference. Crowell seems to think of the wanting-to-have-a-conscience as an answer to Pippin’s objection, while I believe that inauthentic Dasein can already act in light of norms, so that Pippin’s complaint is misplaced. Wanting-to-have-a-conscience enables Dasein not to act in light of norms but to act in light of norms that are now made one’s *own*. Despite my misgivings about his reading of Being-guilty as Being-responsible, I believe my short analysis of wanting to have a conscience chimes with Haugeland’s concept of “resolute responsibility” (Haugeland 2013, 238–239).

collective conscience. Moreover, in both cases, the resulting individual is not necessarily a radical moral maverick who sets out to disturb or replace the norms of his or her society. Rather, in both cases, a different relation to these norms is attained or cultivated. One important difference is that the inauthentic, irresolute Dasein of everydayness can experience its own, personal feeling of guilt and bad conscience⁵³ and is not tethered to some collective form of self-flagellation, as is the case with Nietzsche's description of the single person who still belongs to the morality of customs. Individuation in the form of authentic Dasein is not so much psychological as existential. While Nietzsche is interested in tracking the psychological transformations that led to the emergence of the sovereign individual, Heidegger seeks to identify the existential structures and phenomena that attest to the possibility of authenticity.

While I cannot enter here into a more detailed analysis of Heidegger's concept of authenticity, it nevertheless seems, as I have said, that conceiving of wanting to have a conscience as a necessary condition for responsibility is placing this important concept on rather shaky grounds. To bite the bullet in this way and subscribe to Heidegger's view of responsibility as conditioned by wanting to have a conscience would require an especially sturdy set of teeth, for the resulting view would clash rather violently with our normal ascriptions of responsibility and guilt, where these do not presuppose authenticity. I conclude that Heidegger's idea of wanting-to-have-a-conscience remains at the end of the day rather problematic; it is not clear how it is supposed to supply the necessary condition for responsibility/answerability.

To close this discussion, I wish to consider an important issue that arises specifically with respect to factual *moral* guilt. As we saw, Being-guilty supplies the necessary condition of guilt in general, but to be able to be factually *morally* guilty, Dasein, on the interpretation provided in this chapter, must have first projected itself

⁵³ Though perhaps not fully justifiably, if indeed genuine guilt is possible only when one becomes resolute.

onto the possibility of *morality*. We thus seem to be confronted with yet another ontic condition of moral guilt, one the obtaining of which cannot be assumed in advance for any specific Dasein. This, too, seems to threaten to get people off the hook: cases of Dasein who did not project themselves upon the possibility of Being of morality could not be held responsible for failing to meet morality's demands since these demands would not apply to them.

This critical point must, however, be balanced by an idea that Heidegger makes quite explicit, namely, that since Dasein as Being-in-the-world is essentially with others, it follows that whatever possibility-of-Being it projects itself onto will involve norms that in some way or at some level concern others. As Heidegger puts it, Dasein, "on the null basis of its null projection . . . has, in Being with Others, already become guilty towards them" (BT 334). Nevertheless, this Being-guilty toward others seems to receive content only through the various ontic possibilities onto which Dasein projects itself. In other words, it seems that there is in Heidegger's view no specific possibility of Being for Dasein that cannot be outstripped and that concerns one's relation to others qua others, that is, qua other human beings (rather than qua one's children, spouse, fellow citizens, clients, etc.). There is no *moral* possibility of Being that belongs to Dasein as such. Kant, from a Heideggerian point of view, was wrong in thinking one can articulate a world-less moral "algorithm" that possesses authority with respect to every Dasein as such. But Heidegger, arguably, was wrong in denying the inalienable moral concern we have for others qua human beings. Heidegger's view is thus close to Nietzsche's or Rée's: morality is a specific historical development and does not characterize the Being of the human being as such.

VI. Conclusion

In conclusion, to return to my opening remarks, in Heidegger's analysis of Being-guilty, we can see how Heidegger provides us with

a synthesis of sorts of the two approaches considered—the metaphysical and the naturalist. On the one hand, the idea of *causa sui* is rejected (with the naturalists), but on the other, the notion that the human being is guilty as such, and in a manner that makes guilt possible, is retained (with the metaphysicians), since, for Heidegger, our Being-guilty—a condition of possibility of guilt—is constituted by our *not* being *causa sui*. Where, however, does the discussion presented in the foregoing chapters leave us? Is guilt at the end of the day justified or not? In the concluding chapter, I will offer an answer to this question.

Conclusion

Where does the foregoing discussion leave us? We might agree with Nietzsche that the concept of *causa sui* is indeed untenable and agree with Heidegger that not being *causa sui* is a necessary condition for guilt. But it now might seem we are facing an aporia of sorts: on the one hand, we have the Nietzschean argument that one must be *causa sui* for guilt to be justified, but on the other hand, we have the Heideggerian argument that *not* being *causa sui* is a necessary condition for guilt to be possible. How can we make sense of this predicament? Moreover, where do we now stand with respect to the justifiability of guilt?

To show why this is merely an apparent aporia, we must first distinguish the scope of the *sui* or “self” at stake in each case and see that the “self” that Nietzsche is concerned with is not the “self” that Heidegger has in mind. Nietzsche understands the “self” at issue here as a *physio-psychological* entity (cf. BGE 12, 19): an array of embodied psychological forces such as drives, instincts, feelings, and emotions organized (to some degree) in hierarchical relations of command and obedience. To be *causa sui* in this way is to be able to determine by oneself one’s entire physio-psychological being and thus determine one’s actions purely by oneself. Since, according to Nietzsche, this is absurd, and since this is a necessary element in the justification of guilt, guilt cannot be justified. On the other hand, the self that Heidegger has in mind is one that is to be understood *existentially* (e.g., BT 152–153), specifically, as an entity whose Being is *care* and does not have the kind of Being of presence-at-hand that the entities posited by science in general and

psychology specifically possess. What this means is that the self, for Heidegger, cannot be reduced to its physio-psychology. As we saw, for Heidegger, our self is essentially governed by normativity. Consequently, to be *causa sui* must mean to be able to determine by oneself those normative demands. According to Heidegger, this is something that lies beyond our power as the kind of beings that we are, and yet this is precisely what makes guilt possible. Who is right about the self, then?

Fortunately, we do not have to settle this question here, for even if Heidegger is right that the self is essentially normative and that therefore not being *causa sui* in his sense is a necessary condition for guilt, Nietzsche could still make the argument that for *factual* guilt to be *justifiably* attributable (to oneself by oneself or by others), one must be *causa sui* in *his* sense of the term. Nietzsche can thus insist that the fact that we are not *causa sui* in *his* sense is what precludes factual guilt from ever being justified. This is not, however, a very satisfying result, for while it does indeed show there is no contradiction between Nietzsche's and Heidegger's views, it leaves the question of justification answered in the negative and thus renders the Heideggerian position almost utterly empty and idle; Being-guilty is left as a condition of possibility of something that is never really justified.

A possibly more appealing conclusion to our discussion—one that I will espouse here—would be to reject Nietzsche's claim that factual guilt is unjustified. This would be to hold that Nietzsche is right that we cannot be *causa sui* for any specific action, which is why he is right to hold that people do not deserve punishment (e.g., HH 105; TI, Errors, 7) but wrong to think that guilt, the experience or feeling of guilt, is unjustified. Specifically, I want to suggest that the problem with Nietzsche's view is that he has the following, implicit, argument:

- (1) Guilt is a form of self-punishment.
- (2) Punishment (of whatever kind) is justified only if one is morally responsible for the wrong one is punished for.

(3) No one is morally responsible (we are not *causa sui*).

(4) Hence, guilt is unjustified.

But the argument is problematic. Premise 1, I wish to contend, is unjustified; to use Nietzsche's own words, it "is interpretation, not text" (BGE 22). Now, as far as I know, Nietzsche does not state premise 1 explicitly, but there is reason to think that this is his view. To start, already in HH 39, as we know, Nietzsche rejects the idea that the pain of guilt is justified, for he thinks that lack of free will means that no one is responsible for his deeds and that moral judgment, even "when the individual judges himself" (HH 39), is unjust. But in the same work, and on the basis of the same consideration (lack of responsibility), he rejects the idea that people deserve *punishment* (HH 105). This strong connection in Nietzsche's mind between guilt and punishment persists all the way to the end of his career. Thus, in *Twilight*, he claims that "people were considered 'free' so that they could be judged and punished—so that they could be *guilty* [*schuldig*]" (TI, Errors, 7, see also D 78). Finally, and most tellingly, the *Genealogy* itself displays this conception of guilt as a form of punishment in the most vivid way. First, guilt in its basic form for Nietzsche, of course, is cruelty directed at one-self. But second, the *Genealogy*'s genealogy of guilt is itself full of descriptions of guilt as a form of self-inflicted *punitive violence*: it is described as a form of "self-crucifixion," "self-violation," self-"laceration" (GM II:23), self-"torment" and self-"torture" (GM II:6, 16, 22). And given that crucifixion, torture, and so on, are forms of punishment, it becomes quite clear how in Nietzsche's mind, not only are the guilty—according to the morality he rejects—punished in various ways by an external authority of some kind or another, but also the guilt they themselves experience is *itself* a form of punishment. Since, however, nobody deserves to be punished in any way, guilt, too, Nietzsche thinks, is unjustified.

What, however, are Nietzsche's reasons for interpreting guilt as self-punishment? His account of guilt as resulting from the

internalization of cruelty cannot be the *grounds* for such an interpretation of guilt, for it is supposed to be the *explanation* that sheds light on guilt's unpleasant nature, which must therefore be independently assumed to be a form of self-inflicted punishment (Nietzsche's genealogy cannot create its explanandum as it goes). But what reason is there to interpret guilt as self-torture and so on? Nietzsche, of course, bases his interpretation of guilt on the unpleasant qualitative nature of the feeling of guilt. This is indeed undisputable "text." But—again—why describe this feeling as a form of self-directed punishment (crucifixion, laceration, torture)? The suspicion arises that Nietzsche, despite his aspiration to overcome all the shadows of the dead God (GS 108), is still trapped within the vengeful, punitive, religious-moralistic web of concepts.

I would like here to sketch a different way of thinking about guilt. According to this alternative view, two distinct elements in the experience of guilt should be distinguished. *First*, there is the basic painful feeling of guilt which arises upon the recognition that one has done something that goes against one's values, commitments, and so on. This first element does not depend on any assumption regarding free will or responsibility in the robust, moral sense.¹ Importantly, there is no condemning *judgment* at this level. Here there is only the pain of the recognized contradiction between the values or norms one adheres to, on the one hand, and the nature of the action one has performed, on the other. The pain of guilt that is felt here is that of the tear or split that forms when *who* one takes oneself to be (the norms, values, and so on, that one is committed to) and *what* one has done stand at odds with each other. The basic feeling of guilt then should be understood as a painful, emotional expression of one's attachments to certain values and is to that extent perfectly rational and justified: it is rational to feel pain when

¹ Gabriele Taylor, for one, believes that guilt as such does not require responsibility in a more robust sense than causal: "Causal responsibility is the type that is sufficient for guilt, and that much is also necessary" (Taylor 1985, 91).

one's actions go against what one cares about or is committed to.² This is not to deny that when we *are* morally responsible (when we performed our action intentionally and without coercion, etc.), we feel the pain of guilt much more strongly, but not—I suggest—because we punish ourselves more severely on account of being morally responsible and thus more deserving of our own wrath but because when the action is more genuinely ours (we knew what we were doing, etc.), we experience a stronger practical contradiction between what we value and adhere to and what we did. The more we can conceive of the action as ours, the greater the contradiction, the greater the pain. Furthermore, this rift, it is important to emphasize, is not the intentional *object* of guilt; one does not feel guilty for creating this division between one's values and action. Rather, the guilt is focused either at the action or at the self who performed the action, depending on the severity of the action, where the more severe the transgression, the more the focus is directed at the self.

Then, in addition, as a *second* element, there is, possibly, a *judging* condemnation of oneself expressed in such thoughts as “why did I do that?” or “I could have acted differently!”—the latter thought especially made possible for us with the invention of the idea of free will and the use to which it has been put, namely, as Nietzsche insightfully perceives, to justify punishment and guilt. It is this second element that can be experienced and can be plausibly interpreted as a form of *punishment* or self-laceration. In my view, then, Nietzsche *reduces* guilt to this second element, which, plausibly in my view, he considers unjustified given that freedom of will is an illusion.

The reason I find this alternative view of guilt just sketched attractive is, first, that it combines the insights of Nietzsche and

² My view is thus similar to that expressed by Thomas E. Hill Jr., though I take guilt to extend beyond the moral sphere: “The pain [of guilt] may be a natural human reflection of our recognition that we have failed to show respect for the moral standards that we recognize as authoritative. A bad conscience hurts because, sometimes in spite of ourselves, we care about whether we make our moral judgments with due care and live by them” (Hill 2002, 249).

Heidegger without, however, accepting Nietzsche's claim about the unjustifiability of guilt. We can agree with Nietzsche that self-torture is unjustified but maintain that the self-accusing thoughts do not constitute guilt in its entirety and that the negative emotion at the basis of guilt *is* justified insofar as it expresses our caring about certain norms or values that we subscribe to. The realization that self-accusation is unjustified would perhaps help dull some of the sting of conscience but would not render the experience in its entirety unjustified.

Furthermore, this two-tiered view, I believe, corresponds to experience given that it often happens that guilt is experienced *without* any form of self-flagellation. This typically is the case when the transgression is relatively mild and so the guilt is quick to dissipate. For example, I might feel guilty for showing up late to pick up my kids from school without internally berating and flagellating myself with such thoughts as "but I could have done things differently!" If this is correct, then one can feel guilt over what one has done or didn't do without experiencing the additional self-punishment Nietzsche describes. Guilt does not essentially involve a moment of self-laceration. This would be guilt in the "thin" sense.

How, however, is guilt in this "thin" sense to be distinguished in my view from regret? On some views (e.g., Wallace 2017), regret is a backward-looking painful emotion that concerns an event that has harmed a person or a project or an ideal that we cherish. Importantly, regret concerns not only harm that merely happened to affect our valued objects or persons but also harm that we ourselves have brought about through our own action (which is called, after Williams 1993a, "agent regret"), though involuntarily and in a manner that is not subject to reproach or criticism. This sounds very close to my view of "thin" guilt.

In response, I wish first to draw a distinction. On the one hand is a *general* concept of regret where the term refers merely to the sorrowful recognition that something undesired has occurred, whether as a result of one's own doing (agent regret) or not,

something one would have prevented from happening if one could (all things considered). This kind of regret could also be experienced in a case of guilt in my sense: we can regret an action for which we feel guilty. On the other hand, there is regret in a more *specific* sense, which merely concerns one's own *prudential* or *cognitive* failures. Thus, when I play the chess move that loses the game, when I invest my money in a losing stock, when I say something at a party which I later realize will cost me my promotion—in such cases, unless, of course, we add additional features to their description—it seems that regret (or agent regret) in this *specific* sense, rather than guilt, would be expected and justified (though I regret playing the move, I don't feel guilty for it). Guilt, on the other hand, I submit, is experienced when we fail to live up to certain demands that we take to apply to ourselves but where the failure is not specifically one of prudence or cognition. Guilt would occur in cases where more substantive values or norms of intrinsic importance for us are at issue, cases that do not merely concern one's prudence, intelligence, or capacity for judgment.³

³ Remorse, in my view, is the regret (in the general sense) one feels when one is guilty for an *immoral* action for which one regards oneself as responsible (in a compatibilistic sense: one knew what one was doing, etc.).

Bibliography

Works by Kant

- Critique of Practical Reason*. Translated by Mary J. Gregor. In *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Cambridge: Cambridge University Press, 1996.
- Critique of Pure Reason*. Translated by Paul Guyer and Allen Wood. Cambridge: Cambridge University Press, 1998.
- Groundwork of the Metaphysics of Morals*. Translated by Mary J. Gregor. In *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Cambridge: Cambridge University Press, 1996.
- The Metaphysics of Morals*. Translated by Mary J. Gregor. In *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Cambridge: Cambridge University Press, 1996.
- Religion within the Boundaries of Mere Reason*. Translated by Allen Wood and George di Giovanni. In *Religion within the Boundaries of Mere Reason and Other Writings*. Cambridge: Cambridge University Press, 1998.

Works by Nietzsche

- Beyond Good and Evil*. Translated by Walter Kaufmann. New York: Vintage, 1966.
- The Birth of Tragedy*. Translated by Ronald Speirs. In *The Birth of Tragedy and Other Writings*. Cambridge: Cambridge University Press, 1999.
- Daybreak*. Translated by John R. Hollingdale. Cambridge: Cambridge University Press, 1997.
- Ecce Homo*. Translated by Walter Kaufmann. In *On the Genealogy of Morals and Ecce Homo*. New York: Vintage, 1967.
- The Gay Science*. Translated by Walter Kaufmann. New York: Vintage, 1974.
- The Gay Science*. Translated by Josefine Nauckhoff. Cambridge: Cambridge University Press, 2001.
- Human, All Too Human*. Translated by John R. Hollingdale. Cambridge: Cambridge University Press, 1986.
- Kritische Studienausgabe*. Edited by G. Colli and M. Montinari. Berlin: De Gruyter, 1999.

- On the Genealogy of Morals*. Translated by Walter Kaufmann and John R. Hollingdale. In *On the Genealogy of Morals and Ecce Homo*. New York: Vintage, 1967.
- On the Genealogy of Morals*. Translated by Carol Diethe. Cambridge: Cambridge University Press, 2006.
- Philosophy in the Tragic Age of the Greeks*. Translated by Marianne Cowan. Washington, DC: Gateway, 1998.
- Twilight of the Idols*. Translated by Judith Norman. In *The Anti-Christ, Ecce Homo, Twilight of the Idols: And Other Writings*. Cambridge: Cambridge University Press, 2005.
- The Wanderer and His Shadow*. Translated by John R. Hollingdale. In *Human, All Too Human*. Cambridge: Cambridge University Press, 1986.
- The Will to Power*. Translated by Walter Kaufmann and John R. Hollingdale. New York: Vintage, 1968.

Works by Schopenhauer

- On the Basis of Morality*. Translated by E. F. J. Payne. Indianapolis: Hackett, 1995.
- Prize Essay on the Freedom of the Will*. Translated by E. F. J. Payne. Cambridge: Cambridge University Press, 1999.
- The World as Will and Representation*, Vols. I and II. Translated by E. F. J. Payne. New York: Dover, 1969.

Other Works

- Abbey, Ruth. *Nietzsche's Middle Period*. Oxford: Oxford University Press, 2000.
- Acampora, Christa Davis. "On Sovereignty and Overhumanity: Why It Matters How We Read Nietzsche's *Genealogy* II: 2." In *Nietzsche's On the Genealogy of Morals: Critical Essays*, edited by Christa Davis Acampora, 147–161. New York: Rowman & Littlefield, 2006.
- Allison, Henry E. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1990.
- Ansell-Pearson, Keith. "Nietzsche: A Radical Challenge to Political Theory." *Radical Philosophy* 54 (Spring 1990): 10–18.
- Atwell, John E. *Schopenhauer: The Human Character*. Philadelphia: Temple University Press, 1990.
- Beiser, Frederick C. *Weltschmerz: Pessimism in German Philosophy, 1860–1900*. Oxford: Oxford University Press, 2016.
- Bernstein, Richard J. *Radical Evil: A Philosophical Investigation*. Cambridge: Polity, 2002.

- Blattner, William. "Life Is Not Literature." In *The Many Faces of Time*, edited by J. B. Brough and L. Embree, 187–201. Dordrecht: Kluwer Academic, 2000.
- Blattner, William. "Essential Guilt and Transcendental Conscience." In *Heidegger, Authenticity, and the Self: Themes from Division Two of Being and Time*, edited by Denis McManus, 116–134. London: Routledge, 2015.
- Braver, Lee. *Heidegger*. Cambridge: Polity, 2014.
- Brehm, Alfred Edmund. *Illustriertes Thierleben: Eine allgemeine Kunde des Thierreichs*. 6 vols. Hildburghausen: Verlag der Bibliographischen Instituts, 1864–1869.
- Cokelet, Brad, and Corey J. Maley, eds. *The Moral Psychology of Guilt (Moral Psychology of the Emotions)*. London: Rowman & Littlefield, 2019.
- Crowell, Steven. "Conscience and Reason: Heidegger and the Grounds of Intentionality." In *Transcendental Heidegger*, edited by Steven Crowell and Jeff Malpas, 43–62. Stanford, CA: Stanford University Press, 2007.
- Deleuze, Gilles. *Nietzsche and Philosophy*. Translated by Hugh Tomlinson. London: Continuum, 2006.
- DiCenso, James J. *Kant's Religion within the Boundaries of Mere Reason: A Commentary*. Cambridge: Cambridge University Press, 2012.
- D'Iorio, Paolo. *Nietzsche's Journey to Sorrento: Genesis of the Philosophy of the Free Spirit*. Translated by Sylvia Mae Gorelick. Chicago: University of Chicago Press, 2016.
- Dreyfus, Hubert L. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press, 1992.
- Dreyfus, Hubert L. *Being-in-the-World: A Commentary on Division 1 of Heidegger's "Being and Time"*. Cambridge, MA: MIT Press, 1993.
- Duffy, James. "Homer's Conception of Fate." *Classical Journal* 42, no. 8 (May 1947): 477–485.
- Elgat, Guy. *Nietzsche's Psychology of Ressentiment: Revenge and Justice in On the Genealogy of Morals*. New York: Routledge, 2017.
- Elgat, Guy. "Heidegger on Guilt: Reconstructing the Transcendental Argument in *Being and Time*." *European Journal of Philosophy* 28, no. 4 (2020): 911–925.
- Engstrom, Stephen. "The Inner Freedom of Virtue." In *Kant's Metaphysics of Morals: Interpretative Essays*, edited by Mark Timmons, 289–315. New York: Oxford University Press, 2002.
- Fischer, Martin John, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998.
- Franco, Paul. *Nietzsche's Enlightenment: The Free-Spirit Trilogy of the Middle Period*. Chicago: University of Chicago Press, 2011.
- Gardiner, Patrick. *Schopenhauer*. Bristol: Thoemmes, 1963.
- Gasdaglis, Katherine. "Moral Regret and Moral Feeling(s)." *Inquiry: An Interdisciplinary Journal of Philosophy* (March 17, 2019). <https://www.tandfonline.com/doi/abs/10.1080/0020174X.2019.1592701>.

- Gemes, Ken. "Nietzsche on Free Will, Autonomy, and the Sovereign Individual." In *Nietzsche on Freedom and Autonomy*, edited by Ken Gemes and Simon May, 33–49. Oxford: Oxford University Press, 2009.
- Golob, Sacha. *Heidegger on Concepts, Freedom and Normativity*. Cambridge: Cambridge University Press, 2014.
- Graeber, David. *Debt: The First 5000 Years*. New York: Melville House, 2011.
- Guyer, Paul. "Problems with Freedom: Kant's Argument in *Groundwork III* and Its Subsequent Emendation." In *Kant's "Groundwork of the Metaphysics of Morals": A Critical Guide*, edited by Jens Timmermann, 176–202. Cambridge: Cambridge University Press, 2009.
- Harman, Gilbert, and Corey J. Maley. "The Feeling of Guilt." In *The Moral Psychology of Guilt (Moral Psychology of the Emotions)*, edited by Brad Cokelet and Corey J. Maley, 13–36. London: Rowman & Littlefield, 2019.
- Hatab, Lawrence. *A Nietzschean Defense of Democracy*. La Salle, IL: Open Court, 1995.
- Haugeland, John. *Dasein Disclosed: John Haugeland's Heidegger*, edited by Joseph Rouse. Cambridge, MA: Harvard University Press, 2013.
- Heidegger, Martin. *Being and Time*. Translated by John Macquarrie and Edward Robinson. New York: Harper & Row, 1962.
- Heidegger, Martin. *Schelling's Treatise on the Essence of Human Freedom*. Translated by Joan Stambaugh. Athens: Ohio University Press, 1985.
- Hill, Thomas E., Jr. "Punishment, Conscience, and Moral Worth." In *Kant's Metaphysics of Morals: Interpretative Essays*, edited by Mark Timmons, 233–253. New York: Oxford University Press, 2002.
- Irlenborn, Bernd. "Das Problem der Schuld in Heidegger's *Sein und Zeit*." *Perspektiven der Philosophie: Neues Jahrbuch* 30 (2004): 189–208.
- Jacquette, Dale. *The Philosophy of Schopenhauer*. McGill–Queen's University Press, 2005.
- Janaway, Christopher. *Self and World in Schopenhauer's Philosophy*. Oxford: Oxford University Press, 1989.
- Janaway, Christopher. *Beyond Selflessness: Reading Nietzsche's Genealogy*. Oxford: Oxford University Press, 2007.
- Katchadourian, Herant. *Guilt: The Bite of Conscience*. Stanford, CA: Stanford University Press, 2011.
- Kierkegaard, Søren. *Either/Or: A Fragment of Life*. Translated by Alastair Hannay. London: Penguin, 1992.
- Kleingeld, Pauline. "Moral Consciousness and the 'Fact of Reason.'" In *Kant's Critique of Practical Reason: A Critical Guide*, edited by Andrew Reath and Jens Timmermann, 55–72. Cambridge: Cambridge University Press, 2010.
- Korsgaard, Christine. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

- Kosch, Michelle. "Idealism and Freedom in Schelling's *Freiheitsschrift*." In *Interpreting Schelling: Critical Essays*, edited by Lara Ostarcic, 145–159. Cambridge: Cambridge University Press, 2014.
- Leiter, Brian. *Nietzsche on Morality*. London: Routledge, 2002.
- Leiter, Brian. "Who Is the 'Sovereign Individual'? Nietzsche on Freedom." In *Cambridge Critical Guide to Nietzsche's On the Genealogy of Morality*, edited by Simon May, 101–119. Cambridge: Cambridge University Press, 2011.
- Long, A. Arthur. "Hellenistic Ethics and Philosophical Power." In *Hellenistic History and Culture*, edited by Peter Green. Berkeley: University of California Press, 1993. <https://publishing.cdlib.org/ucpressebooks/view?docId=ft0000035f&chunk.id=ch5&toc.depth=1&toc.id=ch5&brand=ucpress>.
- Marx, Werner. *The Philosophy of F. W. J. Schelling: History, System, and Freedom*. Translated by Thomas Nenon. Bloomington: Indiana University Press, 1984.
- Matar, Anat. *No Moral Ground: On the Poverty of Ethics* [in Hebrew]. Bnei-Brak: Hakibbutz Hameuchad, 2019.
- McDowell, John. "Virtue and Reason." In John McDowell, *Mind, Value, and Reality*, 50–73. Cambridge, MA: Harvard University Press, 1998.
- McKenna, Michael and Derek Pereboom. *Free Will*. Routledge Contemporary Introductions to Philosophy. London: Routledge, 2016.
- McManus, Denis, ed. *Heidegger, Authenticity, and the Self: Themes from Division Two of Being and Time*. London: Routledge, 2015.
- McPherran, L. Mark. "Socratic Theology and Piety." In *The Bloomsbury Companion to Socrates*, edited by John Bussanich and Nicholas D. S. Smith, 257–275. London: Bloomsbury, 2013.
- Miyasaki, Donovan. "Nietzsche contra Freud on Bad Conscience." *Nietzsche-Studien* 39, no. 1 (2010): 434–454.
- Morrisson, Iain. "Nietzsche's Genealogy of Morality in the Human, All Too Human Series." *British Journal for the History of Philosophy* 11, no. 4 (2003): 657–672.
- Pasternack, Lawrence. *Kant on Religion within the Boundaries of Mere Reason*. New York: Routledge, 2014.
- Pattison, George. "Death, Guilt, and Nothingness in Luther, Kierkegaard, and *Being and Time*." In *Heidegger, Authenticity, and the Self: Themes from Division Two of Being and Time*, edited by Denis McManus, 56–71. London: Routledge, 2015.
- Pippin, Robert. "On Being Anti-Cartesian: Hegel, Heidegger, Subjectivity, and Sociality." In Robert Pippin, *Idealism and Modernism: Hegelian Variations*, 375–394. Cambridge: Cambridge University Press, 1997.
- Plato. *Gorgias*. Translated by Donald J. Zeyl. Indianapolis: Hackett, 1987.

- Plato. *Completed Works*. Edited by John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett, 1997.
- Polt, Richard. *Heidegger: An Introduction*. Ithaca, NY: Cornell University Press, 1997.
- Prinz, Jesse J. *Beyond Human Nature*. New York: Norton, 2012.
- Raffoul, François. *The Origins of Responsibility*. Bloomington: Indiana University Press, 2010.
- Rée, Paul. *Basic Writings*. Edited and translated by Robin Small. Champaign: University of Illinois Press, 2003.
- Rée, Paul. *Gesammelte Werke 1875–1885*. Edited with introduction and commentary by Hubert Treiber. Berlin: De Gruyter, 2004.
- Reginster, Bernard. "What Is the Structure of *Genealogy of Morality* II?" *Inquiry: An Interdisciplinary Journal of Philosophy* 61, no. 1 (2018): 1–20.
- Richardson, John. "Nietzsche's Freedoms." In *Nietzsche on Freedom and Autonomy*, edited by Ken Gemes and Simon May, 127–149. Oxford: Oxford University Press, 2009.
- Ridley, Aaron. *Nietzsche's Conscience*. Ithaca, NY: Cornell University Press, 1998.
- Risse, Mathias. "On God and Guilt: A Reply to Ridley." *Journal of Nietzsche Studies* 29, (2005): 46–53.
- Rukgaber, Matthew. "The 'Sovereign Individual' and the 'Ascetic Ideal': On a Perennial Misreading of the Second Essay of Nietzsche's *On the Genealogy of Morality*." *Journal of Nietzsche Studies* 43, no. 2 (Autumn 2012): 213–239.
- Saunders, Joe. "Kant and Degrees of Responsibility." *Journal of Applied Philosophy* 36, no. 1 (2019): 137–154.
- Schelling, F. W. J. *Philosophical Investigations into the Essence of Human Freedom*. Translated by Jeff Love and Johannes Schmidt. Albany: State University of New York Press, 2006.
- Small, Robin. *Nietzsche and Rée: A Star Friendship*. Oxford: Oxford University Press, 2005.
- Snow, Dale E. *Schelling and the End of Idealism*. Albany: State University of New York Press, 1996.
- Sorabji, Richard. *Moral Conscience through the Ages: Fifth Century BCE to the Present*. Chicago: University of Chicago Press, 2018.
- Strawson, Galen. "The Impossibility of Moral Responsibility." In *Free Will*, edited by Gary Watson, 212–228. Oxford Readings in Philosophy. Oxford: Oxford University Press, 2003.
- Sussman, David. "Review: *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Apriority of History*." *Notre Dame Philosophical Reviews: An Electronic Journal* (July 19, 2010). <http://ndpr.nd.edu/news/kant-s-theory-of-evil-an-essay-on-the-dangers-of-self-love-and-the-apriority-of-history/>.

- Tanzer, Mark Basil. *Heidegger: Decisionism and Quietism*. New York: Humanity Books, 2002.
- Taylor, Gabriele. *Pride, Shame and Guilt: Emotions of Self-Assessment*. Oxford: Oxford University Press, 1985.
- Van Creveld, Martin. *Conscience: A Biography*. London: Reaktion, 2015.
- Versteeg, Russ. *Law in Ancient Egypt*. Durham, NC: Carolina Academic Press, 2002.
- Vlastos, Gregory. *Socrates, Ironist and Moral Philosopher*. Ithaca, NY: Cornell University Press, 1991.
- Wallace, R. J. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.
- Wallace, R. J. *The View from Here*. Oxford: Oxford University Press, 2017.
- Wicks, Robert. *Schopenhauer*. Malden, MA: Wiley-Blackwell, 2008.
- Williams, Bernard. "Moral Luck." In *Moral Luck*, edited by Daniel Statman, 35–55. Albany: State University of New York Press, 1993a.
- Williams, Bernard. *Shame and Necessity*. Berkeley: University of California Press, 1993b.
- Wolf, Susan. *Freedom within Reason*. Oxford: Oxford University Press, 1990.
- Wood, Allen W. *Kant's Moral Religion*. Ithaca, NY: Cornell University Press, 1970.
- Wood, Allen W. "Kant's Compatibilism." In *Self and Nature in Kant's Philosophy*, edited by Allen W. Wood, 73–101. Ithaca, NY: Cornell University Press, 1984.
- Young, Julian. *Willing and Unwilling: A Study in the Philosophy of Arthur Schopenhauer*. Dordrecht: Martinus Nijhoff, 1987.
- Zamosc, Gabriel. "The Relation between Sovereignty and Guilt in Nietzsche's *Genealogy*." *European Journal of Philosophy* 20, no. 1 (2012): 107–142.
- Žižek, Slavoj. *The Indivisible Remainder: An Essay on Schelling and Related Matters*. New York: Verso, 1996.
- Žižek, Slavoj. *The Sublime Object of Ideology*. New York: Verso, 2009.
- Zupančič, Alenka. *Ethics of the Real: Kant, Lacan*. New York: Verso, 2000.

Index

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.

- Abbey, Ruth, 165
 Acampora, Christa Davis, 196
 accountability, accountable, 22–23,
 108–9, 111, 113–14, 251–52,
 254–55. *See also* blameworthy;
 responsibility
 active and reactive in Nietzsche,
 177–78, 179–81, 181n.15,
 182n.16, 198, 231–32, 235–36
 Allison, Henry E., 34–35, 42n.23,
 49–51, 56, 56n.34, 57–58
 Ansell-Pearson, Keith, 196
 ascetic ideal, 181–82, 231, 234–35,
 235n.55
 Atwell, John E., 99–100, 102n.9, 114,
 115–16, 174
 being-guilty (in Heidegger), 7–8,
 241–44, 245–47, 246n.4, 247n.5,
 249–52, 251n.10, 252n.14,
 252–53n.15, 253n.16, 253–55,
 254n.19, 256, 265–66n.34, 270,
 271, 272–73, 275–77, 277n.41,
 278–83, 280n.45, 284–85, 287,
 288, 289, 292–93, 293n.52, 294–
 96. *See also* guilt: ontological
 Beiser, Frederick C., 110n.12
 Bernstein, Richard J., 41–42n.22,
 43n.26, 46n.29, 79n.11
 blameworthy, 128, 128n.15, 131n.16,
 132–33, 134, 137–38, 144–45,
 157. *See also* accountability,
 accountable; responsibility
 Blattner, William, 264n.28,
 265–66n.34
 Braver, Lee, 277n.41
 Brehm, Alfred Edmund, 126,
 126n.12
 causa sui, 3–4, 7–9
 Heidegger and, 7, 8, 241–44,
 248n.6, 248–49, 250–51, 253–
 54, 276–78, 276n.40, 277n.41,
 279–80, 295–96, 297–98, 299
 in Kant, 4, 16, 54–55, 56–57
 Nietzsche's criticism of, 5–6, 8,
 163–64, 171–74, 215, 225, 230,
 297–98, 299
 Rée and, 5–6
 in Schelling, 4–5, 43n.26, 59, 64–
 65, 77–78
 in Schopenhauer, 4–5, 43n.26, 83,
 111, 112, 115–19
 character, intelligible and empirical
 Heidegger on, 248–49
 in Kant, 4–5, 29–30, 42–43, 43n.25,
 48, 55, 56–57, 56n.34, 76
 Nietzsche on, 167–68, 173
 Rée on, 137–38, 148, 148n.29
 in Schopenhauer, 83, 91–100,
 102–3, 102n.9, 104, 108–11,
 109n.10, 110n.12, 112–13,
 113n.13, 114–17
 choice. *See also* intelligible deed
 of being in Schelling, 64–65, 78–
 79, 81–82n.12

- choice (*cont.*)
 of character in Schopenhauer, 4–5,
 110–11, 110n.11, 110n.12, 112–
 14, 115–16, 117
 of fundamental maxim in Kant,
 4, 42–46, 47, 48, 49–50, 51–52,
 53–55, 77–78
 in Kant (*Willkür*), 30–38,
 34n.14, 35n.15, 39, 43–44,
 46–47, 53
- Christianity, Christian(s), 41–42,
 103–4, 175–76, 226–33,
 269–70
- Cokelet, Brad, 8n.5
- compassion, 107, 108, 126, 127, 132,
 143, 144–46, 155–56, 183n.17,
 239–40, 240n.61
- compatibilist, compatibilism, 19n.3,
 28, 47, 75n.9, 196n.26, 241n.1,
 252–53, 290, 303n.3
- conscience, 1–2, 7, 8–10, 11–12,
 301n.2, 301–2
 Heidegger on the conscience, 269;
 wanting to have a conscience,
 287 *passim*, the, 244–45,
 252n.14, 272–73, 281–87, 291,
 291n.50, 293n.52, 293
 Kant on the conscience, 17 *passim*,
 19n.3, 23–26, 28–30, 32–
 33, 38–39
 Nietzsche on bad conscience, 208
passim, 163, 165–66, 168–69,
 169n.3, 170, 175–83, 175–76n.9,
 176n.10, 176n.11, 178n.12,
 182n.16, 183n.17, 203n.30,
 205n.32, 211n.37, 218n.46,
 235n.55, 238n.58
 Rée on bad conscience, 120n.1,
 122, 123–41, 147n.27, 148n.28,
 152n.31
 Schopenhauer on the conscience,
 97–99, 100–3, 102n.9, 106, 107,
 109–11, 112–14
- creditor and debtor, 186–87, 188,
 191, 194–95, 230–31, 239
- Crowell, Steven, 290–91, 291n.50,
 293n.52
- cruelty, 175–76n.9, 177–83, 187,
 189–90, 192, 195–96, 231–33,
 232n.53, 235–37, 236n.57, 239–
 40, 240n.61
- Darwin, Charles, 141, *passim*, 155;
 Darwinian, 122, 123–24, 125–
 27, 126n.11, 139, 143–46
- debt, 186n.21, 187–88, 187n.22,
 191, 191n.24, 192–93, 194–95,
 203n.30, 230–31, 235, 239,
 249, 264–65, 272–73, 277n.41,
 280n.45, 284. *See also* creditor
 and debtor
- Deleuze, Gilles, 181n.15
- determinism, 3–4, 26–27, 39, 47, 53,
 54, 60–61, 62, 73–75, 76, 81–82,
 97–98, 100, 118, 157–59, 160,
 164–65, 217–18n.45
- DiCenso, James, 51n.30
- D'Iorio, Paolo, 122n.6, 216n.42
- Dostoevsky, Fyodor, 285–86
- Dreyfus, Hubert L., 242n.3, 259–60
- drive, 125–28, 126n.11, 129, 130–32,
 138, 141, 142–43, 142n.25, 144–
 45, 155–56, 176–77, 224–25. *See*
also instinct, instinctive
- Duffy, James, 215–17
- Elgat, Guy, 175n.8, 180n.14, 227n.49,
 232n.53, 235n.56, 246n.4,
 252n.12
- empirical guilt. *See* guilt: empirical
 or factual
- Engstrom, Stephen, 16–36
- evil. *See also* sin
 evil in Nietzsche, 165, 176n.10,
 184, 204–5, 208–9, 209n.35,
 225, 226–28, 229–30, 233

- evil in Schelling, 60, 77–78, 79–80
radical evil in Kant, 16, 41–43,
43n.26, 48–51, 58
- factical. *See* guilt: empirical or factical
feeling
- of freedom (in Schelling), 62–
64, 80–81
- of guilt, 1n.1, 10, 11–12, 17–18,
19–20, 38–39, 58, 80–81, 82, 106,
107, 108–9, 111, 112, 113–14,
128, 132–33, 133–34n.17, 134,
134n.18, 136, 139, 142, 143, 144,
145–46, 148n.29, 153–54, 165–
66, 167n.2, 167–70, 174, 175, 177,
180–81, 182–83, 186, 186n.21,
187n.22, 191, 195–96, 201, 210–
11, 212–14, 228, 234–35, 236,
284, 287, 293–94, 298, 299–300
(*see also* frustration of drives/
instincts, guilt)
- Fischer, Martin John, 252–53
- Franco, Paul, 165
- freedom
- intelligible, 3–6, 7, 11, 16, 31n.12,
41–42n.22, 44–45, 47, 49–50,
51–52, 57–58, 59, 74, 78, 80–81,
82, 111, 112, 114n.14, 116,
118, 137–38, 165–66, 167–69,
172–74 (*see also* character,
intelligible and empirical;
intelligible deed)
- leeway, 74, 77n.10, 78, 117–18, 135–
36, 152–53, 157, 164–65, 222–24,
225, 234, 252–53, 254n.18 (*see
also* free will: libertarian)
- practical, 30–33, 31n.12, 34–35, 47
- transcendental, 16–17, 26–30,
31, 31n.12, 33, 40–41, 44–45,
45n.28, 47, 58, 115–16, 117
- free will, 5–6, 8, 12, 60–61, 67, 74,
112–13, 113n.13, 116, 120, 124–
25, 134–36, 137–38, 139, 143,
146, 147, 148, 152–53, 152n.32,
156–58, 160–61, 164–66, 167–
68, 169n.3, 171, 172, 186–87,
195–96, 211, 217, 217–18n.45,
223–24, 232–33, 234, 241n.1,
299, 300–1. *See also* freedom
libertarian, 217, 217–18n.45, 223–
24, 252–53
- Freud, Sigmund, 9–10
- frustration of drives/instincts, 138,
139, 141–46, 142n.25. *See also*
drive; instinct, instinctive
- Gardiner, Patrick, 110n.11
- Gasdaglis, Katherine, 25n.7
- Gemes, Ken, 196n.26
- God, 2–3n.3, 22–23, 22n.5, 24–25,
25n.6, 60–61, 62, 66–70,
67n.6, 68n.7, 71–72, 93n.7,
101, 123–24, 170, 171–72,
194, 226–29, 227n.50, 230–31,
232–33, 269–70, 275–76,
276n.40, 299–300
- Golob, Sacha, 247n.5, 249n.8, 288
- Graeber, David, 238–39, 238n.59,
238n.60
- Greeks, 175–76n.9, 206–7, 211n.37,
215–17, 220
- guilt
- Christian, 229–30 *passim*, 6, 175–
76, 232n.53, 269–70
- empirical or factical, 248–49
passim, 266 *passim*, 7, 11,
40–41, 44, 45–46, 50–51, 52,
54, 57–58, 82, 229–30, 241–42,
244, 264–65, 265n.30, 265n.31,
268n.36, 269n.37, 270, 271,
272–73, 274, 275–77, 277n.41,
279–81, 280n.45, 284–85, 286–
90, 292–93, 294–95, 298
- ontical, 264–65, 266–68, 268n.36,
269, 271, 273, 275–76, 278,
284, 289

- guilt (*cont.*)
 ontological, 270 *passim*, 3–6, 7–8,
 11, 16, 40–42, 43–46, 43–44n.27,
 45n.28, 51–52, 54–55, 57–58, 59,
 78–79, 80–81, 82, 103–4, 110,
 112–13, 118, 168–69, 229–30,
 231–33, 241–42, 247, 248–49,
 248n.7, 249n.8, 250–52, 253–54,
 255, 255n.20, 256, 265–66n.34,
 277n.41, 280n.45, 282–83, 285,
 286–88, 291n.50
 Guyer, Paul, 41n.21, 45n.28
- Harman, Gilbert, 1n.1
 Hatab, Lawrence, 196, 196n.26
 Haugeland, John, 251–52 *passim*,
 251n.11, 252n.13, 252–53n.15,
 253n.16, 253n.17, 254n.18,
 293n.52
 Hegel, G.W.F. 9–10, 9n.7, 63n.3, 72,
 291–92n.51
 Hill, Thomas E., Jr., 19n.3, 301n.2
 Hume, David, 2–3n.3, 151
- imputation, 19n.4, 26n.8, 26–40. *See*
also responsibility: moral
 infinite regress, 4–5, 52–53, 54–55,
 59, 75, 77–78, 117, 173–74
 instinct, instinctive, 126–27,
 127n.14, 141–42, 142n.25, 143,
 144, 155–56, 176–78, 211, 212–
 14, 220, 234
 of freedom, 178–79, 181,
 236, 239–40 (*see also*
 power: will to power)
 intelligible being (in Schelling), 75–
 77, 75n.9
 intelligible deed, 16, 41–42n.22,
 43n.26, 44–45, 45n.28, 47,
 49–50, 57, 58, 59, 77–79,
 80–81, 81–82n.12, 83, 111,
 118, 120–21, 137–38, 165–66,
 174, 272–73
 internalization, 6, 175–76n.9, 177–
 79, 178n.12, 189–90, 231–32,
 236, 299–300. *See also* cruelty
 Irlenborn, Bernd, 280n.45
- Jacquette, Dale, 103–4
 Janaway, Christopher, 88n.1,
 114n.14, 121n.3, 139n.22,
 147n.27, 152n.31, 159, 172,
 176n.10, 231n.52
 justice, 22–23, 29, 164–65, 209–10
 Rée on sense of justice, retributive
 justice, 123–24, 148–54,
 152n.32, 158–59
 Schopenhauer on eternal
 justice, 103–4
- Katchadourian, Herant, 8n.5
 Kaufmann, Walter, 172n.7, 186n.21,
 191n.24, 192n.25
 Kierkegaard, Søren, 269–70, 278–79,
 279n.43, 285–86
 Kleingeld, Pauline, 58n.36
 Korsgaard, Christine, 256–57
 Kosch, Michelle, 75n.9, 77n.10
- Leiter, Brian, 172n.7, 173, 175–
 76n.9, 196, 196n.27, 215n.39,
 217–18n.45
 Long, A. Arthur, 206–7
 Lubbock, John, 129
- Maley, Corey J., 1n.1, 8n.5
 Marx, Werner, 68n.7
 masters, masterly (in Nietzsche)
 184, 190
 master morality, 225–26 (*see also*
 noble, nobleman, nobles)
 Matar, Anat, 25n.6
 McDowell, John, 259n.24
 McKenna, Michael, 74
 McManus, Denis, 251n.9
 McPherran, L. Mark, 208–9, 218

- Miyasaki, Donovan, 10n.8
- moral responsibility. *See*
 accountability, accountable;
 blameworthy; guilt;
 responsibility, moral
- Morrisson, Iain, 183n.18, 185n.19
- naturalism, 120n.2, 140
- noble, nobleman, nobles, 190, 220
- nullity, 247–48, 251, 255, 271–
 72, 274, 277n.41, 277–79,
 278n.42, 281
- ontological guilt. *See* guilt:
 ontological
- original sin, 4, 11, 16, 41–42, 43–44,
 45–46, 52, 54, 57–58, 78–79,
 82, 103–4, 105–6, 111, 112–13,
 231–33, 250, 270
- pantheism, 60–61, 62–64, 66, 67
- Pasternack, Lawrence, 42n.23,
 43n.26, 47–48, 49–50
- Pattison, George, 251n.10
- Pereboom, Derek, 74
- Pippin, Robert, 290–91, 291–92n.51
- Plato, 83–84, 115–16, 168–69, 172,
 173, 206–7, 209n.35, 216n.42
- Platonic, 96–97, 218n.46
- Polt, Richard, 242n.3
- power
 Dasein's power over its own being,
 8–9, 248–49, 250–51, 274–76, 278
 will to power, 177, 181, 236
- Prinz, Jesse J., 135n.19, 236
- punishment, 18n.2, 39, 103–4, 105,
 120, 123–24, 130, 133–34n.17,
 142n.25, 147, 147n.27, 148–54,
 149n.30, 156–57, 158–60, 161–
 62, 164–65, 177–78, 184, 186–
 87, 187n.22, 188n.23, 189–90,
 191, 192, 194–95, 197–98, 200,
 201–3, 205n.32, 206n.33, 210,
 211, 211n.37, 222–23, 225, 229–
 30, 235–36, 236n.57, 239–40,
 270, 285–86, 298, 299–300, 301.
See also justice
- radical evil. *See* evil
- Raffoul, François, 254–55, 254n.19
- Ravizza, Mark, 252–53
- Reginster, Bernard, 10n.8, 175–
 76n.9, 182n.16, 187n.22, 196,
 196n.26, 196n.28, 197–98,
 202–4, 203n.30, 208–9, 219,
 219n.47, 232–33
- regret, 10, 23–24, 145, 154–57, 278–
 79, 302–3. *See also* conscience;
 guilt; remorse
- remorse, 10, 98–99, 101, 120n.1,
 124–25, 135–36, 137–39, 141–
 42, 142n.25, 144–46, 154–56,
 165–66, 211, 303n.3. *See also*
 conscience; guilt; regret
- responsibility, 1–2, 2–3n.3, 3–4, 6, 9,
 12, 16, 25–26, 60, 81–82, 105,
 213n.38, 248, 251–55, 251n.10,
 251n.11, 253n.16, 253n.17,
 254n.18, 254n.19, 287–88, 289–
 90, 292–93, 293n.52, 294. *See*
 also accountability, accountable;
 blameworthy
- causal, 12, 12n.11, 300n.1
- moral, 12, 16, 25–40, 43–44, 45–46,
 47, 52, 53, 54, 54n.31, 63–64, 67,
 76–78, 81, 97–98, 100, 100n.8,
 108–11, 109n.10, 110n.12,
 112, 113–14, 114n.14, 115–16,
 117–18, 120, 124–25, 128n.15,
 135–36, 135n.20, 137–38, 142,
 143, 145, 160, 164–66, 168–69,
 171–72, 172n.6, 173, 178n.12,
 186–87, 187n.22, 195–96, 222–
 23, 225, 253–54, 299, 300–1
 and the sovereign individual, 197–
 204, 203n.30

- ressentiment*, 179–80, 182n.16, 220, 223–25, 226–28, 232n.53, 235–36, 236n.57
- revenge, 195, 210, 220, 225, 226–28, 227n.49, 235–36. See also *ressentiment*
- Richardson, John, 196
- Ridley, Aaron, 177–78, 225–26, 231–32
- Risse, Mathias, 180n.13
- Rukgaber, Matthew, 204n.31, 220
- Sartre, Jean-Paul, 172n.7, 255n.20
- Saunders, Joe, 26n.8, 54n.31
- self-constitution, 49–50, 56–57, 77–79, 77n.10, 79n.11, 80–82, 83, 92–93. See also intelligible deed
- self-deception, 167–68, 227n.49
- sin. See original sin
- slave morality, 226–27; slave revolt in morality, 176n.10, 223–24, 225–28, 232–33
- Small, Robin, 122nn.4–6, 123–24n.9, 126n.13, 142, 142–47nn.25–27
- Snow, Dale E., 78–79
- Socrates, 197, 204, 205–10, 205n.32, 206n.33, 208n.34, 209n.35, 209–10n.36, 218, 218n.46, 220–22, 222n.48
- Sorabji, Richard, 8n.6
- sovereign individual, 6, 175–76, 188, 195–223, 225, 228–29, 237–38, 293–94
- Spinoza, Baruch, 60–61
- Strauss, Richard, 285–86, 285n.48
- Strawson, Galen, 173–74
- Sussman, David, 43n.26
- Tanzer, Mark Basil, 248n.7, 255n.20
- Taylor, Gabriele, 8n.5, 10n.9, 135n.19, 300n.1
- timeless deed. See intelligible deed
- transcendental argument, 3–4, 7, 16–17, 44–45, 59, 78, 81, 83, 91, 108–9, 111, 112, 118–19, 165, 167, 229–30, 244, 249, 251–52, 255, 265–66n.34, 266–82
- unconscious, unconsciously, 69–70, 81–82n.12
- Van Creveld, Martin, 205n.32, 218n.46
- Versteeg, Russ, 239
- Vlastos, Gregory, 209–10n.36
- Wallace, R.J., 145, 215n.40, 252–53, 290, 302
- will. See also freedom; free will
in Kant (*Wille*), 26–27, 31–32, 34–36, 39–40, 55
in Schelling, 4–5, 76; universal will, 4–5, 65–66, 68–70, 71, 72–73, 76–77; creaturely will, 77–78
in Schopenhauer, 85–91, 88n.2, 93–94, 95–97, 98–99, 102–8, 111, 112–14, 113n.13, 114n.14, 115–16
- Williams, Bernard, 10n.9, 135n.19, 211n.37, 217n.43, 302
- Wolf, Susan, 12n.11
- Wood, Allen W., 35–36, 37–38, 42n.23, 109n.10
- Woody, Allen, 285–86
- Young, Julian, 88n.2
- Zamosc, Gabriel, 196, 196n.28, 197–98, 219
- Žižek, Slavoj, 60n.1, 68n.7, 81–82, 81–82n.12
- Zupančič, Alenka, 40n.19